EFFECTS OF FRAME SIZE UPON PROGRAMMED LEARNING OF PSYCHOLOGY BY HIGH SCHOOL STUDENTS

By
JOHN MICHAEL FLYNN

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF EDUCATION

UNIVERSITY OF FLORIDA
1968

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to the members of my supervisory committee for the benefit of their guidance and knowledge throughout my doctoral program. In particular, I wish to thank Dr. Dwain E. Small for his aid and encouragement which began in my undergraduate years and continued throughout graduate school.

Dr. Henry S. Pennypacker has enabled me to develop better insights into objective psychology and has helped me to appreciate the study of learning processes. I feel especially indebted to him.

A special thanks goes to Dr. Vynce A. Hines, who, although he is not on my committee, served as acting chairman for a year, providing excellent advice and guidance.

I am deeply indebted to the chairman of my committee, Dr. Wilson H. Guertin. He, more than any single individual, provided the stimulus which caused me to pursue a doctoral degree in educational psychology. Also, I owe to him the high value I place on objective, quantitative research methods. Further, whatever skills I have developed in research methods are in large part due to his guidance and teaching.

There are, of course, many other individuals on the faculty of the University of Florida to whom I feel deeply obligated, but I will only mention three of these. First, Dr. Charles M. Bridges has been a source of lasting stimulation and knowledge. The many hours I have spent interacting with him have been of tremendous value to my professional development. Second, I am grateful to Dr. J. B. White who

helped me out in several tight spots while I was in graduate school. Also, largely through Dr. White's efforts the Florida Educational Research and Development Council provided much of the funds necessary to conduct the research for this dissertation. Third I wish to thank the late Professor Douglas E. Scates, whom I shall always remember as a great scholar. He provided me with deep insights into the research process and taught me to look critically at research methodology.

The efforts of many people were needed to carry out the research described in this dissertation. Among these people to whom I have a lasting debt are Dr. James Smith and the staff of Nova Schools for providing the research environment, Dean Abraham S. Fischler of Nova University for permitting me to complete this project while I was in his employ, the faculty of Education Center at Nova University for their encouragement and advice, Mrs. Mildred Nitzberg for organizing and coordinating the volunteers, and the volunteers themselves who so generously gave of their time and asked nothing in return. I owe a special debt to Mrs. Edna Lehane who not only typed the manuscript in final form, but who typed several earlier drafts and who aided in the research itself.

The last person I will mention deserves much more than a mere thanks. My wife, Bonnie, is directly responsible for my being in graduate school; I returned to school at her insistence and I was able to remain in school by her sacrifice.

John M. Flynn

TABLE OF CONTENTS

Chapter Page

I. PROBLEM, THEORY AND RELATED RESEARCH 1

Problem Statement 1
Related Research 7
Objectives 12

II. PROCEDURES 13

Sample 13
Materials 14
Design 17
Data Sources 18
Attrition 20

III. INSTRUMENT DEVELOPMENT 35

IV. DATA ANALYSIS 43

Criterion Test Scores 43
Completion Time 51
Error Rate 56
Intercorrelations of Dependent and Independent Variables 58
Attitudes 62

V. DISCUSSION AND CONCLUSIONS 71

APPENDICES 79

REFERENCES 139

LIST OF TABLES

Table Page

1. Completion Status of 166 Students Who Were Selected for the Study 21
2. Students by Sex and Grade Level Completing the Study 23
3. Students by Sex and Version Completing the Study 24
4. Means and Standard Deviations of the Available SCAT Scores (Total) of the Students Who Participated 28
5. Analysis of Variance Summary Table: Participation by Versions for SCAT Total Scores 29
6. Means and Standard Deviations of Affirmative Answers to Ten Psychology Background Questions 30
7. Analysis of Variance Summary Table: Frequency of Affirmative Answers to Psychology Background Questions 31
8. Frequency of Test Frames Corresponding to Response Frames in Each Version 36
9. Item Analysis of the Psychology Achievement Test 38
10. Means and Standard Deviations of the Psychology Achievement Test by Version and Sex 44
11. Analysis of Variance Summary Table Sex X Versions for Psychology Achievement Test Scores 46
12. The Number of Students by Sex and Version for Whom Both the Achievement Test Scores and SCAT Scores Were Available 48
13. Means and Standard Deviations of the SCAT and the Achievement Test 49
14. Analysis of Covariance for the Psychology Achievement Test Scores with SCAT Scores as the Covariate 50

Table Page

15. Means and Standard Deviations of Completion Times in Minutes 51

16. Analysis of Variance for Completion Time 52

17. Difference Between Completion Time Means 53

18. Means and Standard Deviation of Errors Made in the Programmed Materials 57

19. Intercorrelations of the Criterion Variables With Several Independent Variables 59

20. Responses to the Question: "Would you like to read more about psychology?" 63

21. Responses to the Question: "Would you rather read a regular textbook on psychology than have it taught by the materials you have just completed?" 64

22. Responses to the Question: "Did you like the way psychology was taught in the materials you have just completed?" 65

23. Responses to the Question: "Would you like to become a psychologist?" 66

24. Responses to the Question: "If you were requested to participate in a similar study sometime in the future, would you?" 68

25. Responses to the Question: "Do you think that a reward should be given to those who participate?" 69

26. Responses to the Question: "Would you like to see the materials which you read become part of the regular course work at Nova High School?" 69

27. Response to the Question: "Did you learn anything about psychology from participating in the study?" 70

28. Reasons given by students who did not complete the study 80

Table Page

29. Data for Students in Version 1 of the Material 132

30. Data for Students in Version 2 of the Material 133

31. Data for Students in Version 3 of the Material 134

32. Data for Students in Version 4 of the Material 135

33. Data for Students in Version 5 of the Material 136

34. Data for Students in Version 6 of the Material 137

35. Data for Students in Version 7 of the Material 138

CHAPTER I: PROBLEM, THEORY AND RELATED RESEARCH

Problem Statement

One of the obstacles to an effective theory of instruction is the inability to identify the relevant variables involved in instruction -- both the variables centered in the learning situation and the variables centered in the learner. Ideally, educators should be able to specify the optimal instructional method for a given student in a given situation. Of course, they cannot at present do this and probably they will never be able to do it without some qualification. However, it is possible to move in this direction by identifying as many variables as possible which affect the learning situation, and then determining both their effect and the effect of their interactions.

The recent advances in auto-instruction -- particularly in computer-based systems -- affords us the opportunity to experimentally manipulate variables and to determine their effects on learning to a degree which is impossible in conventional learning systems. Unfortunately, while research on auto-instruction has repeatedly shown that such techniques do facilitate learning, few conclusions can be drawn about auto-instruction in general. Frequently, there has been little or no consistency of findings about which variables are the important ones or what the effects of the variables studied are (Silberman, 1962; Schramm, 1964). Consequently, while researchers such as Stolurow have recommended that auto-instructional and conventional methods be alternated in the classroom (1963), there are no guidelines for teachers to follow in implementing auto-instruction.

The purpose of the present study is to move toward identifying and isolating some of the variables which influence programmed learning, with the ultimate hope of giving the teacher guidelines for implementation of these and other techniques and with an eye on the future development of a theory of instruction. The variables involved in any learning situation are numerous and a single study cannot hope to manipulate more than a few of them. For example, Fry (1963) has itemized 212 variables in programmed learning.

The more salient independent variables involved in verbal auto-instructional situations can be roughly divided into five categories with a minimum of overlap. These categories are (1) Structural or format variables, (2) Content variables, (3) Learner-centered variables, (4) Structure-learner interaction variables, and (5) Content-learner interaction variables. The structural and content variables are material-centered and are not dependent upon the learner in any way. The learner-centered variables are likewise independent of the structure and content. The two interaction categories, however, are dependent both upon the learner and upon the materials.

The task of studying all possible combinations of the variables involved in auto-instruction is extremely impractical, if not impossible. Even if we assumed only three critical levels of each variable, the number of different combinations of the variables identified by Fry would be 3^{212} (a number containing 106 digits). Yet, three crucial levels areprobably an underestimation for many of these variables.

Consequently, the number of variables included in a single study has to be greatly delimited.

In the present study the single material-centered variable of frame size in a linear auto-instructional text was manipulated in a one-way design. Here frame size is defined as the amount of material presented between elicitations of responses. While some theorists have maintained that very short frames with small steps between frames will provide optimal learning for all individuals (e.g., Skinner, 1958), there is little empirical evidence to support this view. In fact, Deterline (1967), states that steps which are too small "can actually interfere with the desired learning" (p. 212). This is in agreement with Pressey's position that "...detailed small-step programming may be exceedingly useful -- but surely not for everybody on everything" (Pressey, 1960, p. 503). Several studies have investigated the size of the step between frames (e.g., Coulson and Silberman, 1960, 1961; Shay, 1961; Smith and Moore, 1962; and Kapel, 1965), but there have been very few studies of the amount of material within frames (e.g., Krumboltz and Kiesler, 1965) and these are frequently confounded by additional variables (e.g., Briggs, et al., 1962).

It should be pointed out that frame size cannot be manipulated independently of all other material-centered variables. Obviously, the frequency of response is a function of frame size. If more material is presented between overt responses, it follows that there will be fewer responses per unit time and per unit material. Since feedback and the assumed associated reinforcement are contingent upon responses, the frequency of these variables will also vary with the frame size. To a lesser extent, redundancy of material and cues or prompting will also be in part dependent upon frame size. Thus, although only frame size will be directly manipulated, other variables will also vary.

The rationale for selecting frame size for study is based in large part upon what appears to be a gap between reinforcement theory and the mechanics of auto-instruction. Skinner's auto-instructional techniques grew out of his operant studies of reinforcement with animals. In animal studies, a reasonably constant stimulus situation elicits constant responses which are usually either single discrete acts (e.g., bar pressing by rats), or successive approximations to particular acts. Auto-instruction, however, does not entirely fit the operant paradigm as used in the animal studies, since neither the stimuli nor the responses remain constant throughout the program and strictly speaking are unique for each frame. Ignoring content, there is little question that the act of overtly responding can become an operant under stimulus control; however, the interest in programmed instruction is not the act of responding per se, but the specific content of the response. Hence, if we are to assume that the learning of content is under stimulus control and is reinforced by feedback, we have to conclude that one-trial learning is taking place. That is, we have to assume that a single reinforcement of a specific response is sufficient to maintain that response in the student's repertoire over an indefinite period of time. If this is the case, an additional problem arises with small frames in that the responses will perhaps only be related to the limited stimulus content of that frame (Deterline, 1967).

While one-trial learning is a possible explanation there are other problems with programmed instruction fitting the operant paradigm. For instance, auto-instruction fails to match the operant paradigm because the overt responses are elicited by specific prompts or cues within the frame. This situation is inconsistent with Skinner's position in which the operant responses are emitted rather than elicited. Lumsdaine (1962) has discussed this point and suggests that programmed instruction may better fit contiguity theory than it does reinforcement theory and that the relationship between the cues in the frame and the responses parallels the classical conditioning paradigm. Scandura (1966) suggests that learning principles discovered in the laboratory may have little application in teaching because of the differences in the situations. He points out, for example, that " 'shaping,' the operant technique for modifying overt behavior, has not been adequately specified with respect to meaningful verbal (i.e., symbolic) material" (p. 141). These considerations suggest that operant reinforcement principles cannot be directly applied to auto-instruction, and calls into question the practice of extrapolating from laboratory operant studies to human classroom learning situations.

Even if reinforcement theory can be applied directly to auto-instruction it has not been demonstrated that feedback to the student as to the correctness or incorrectness of his response does, in fact, provide reinforcement to him. If the student's response is incorrect, viewing the correct response is not a parallel situation to the laboratory counterpart where some organism fails to emit a correct response. In the auto-instruction situation, viewing the correct response provides a further cue for a covert or implicit correction of the previous incorrect response. In the laboratory situation an incorrect response usually produces no overt reaction at all from the experimenter of the apparatus (except in avoidance studies).

If the student's response is correct, seeing the correct response does not necessarily constitute reinforcement for the student. Briggs, et al., (1962) point out that for easy frames overt responses and the resultant feedback are not always effective. In a study, reported by Briggs and his colleagues (1962), students who only read frames and did not overtly respond had superior test scores and superior time-efficiency for auto-instructional frames of low difficulty (based on the degree of cueing) over those students who had to overtly respond to the same frames. Licklider (1962) states that on a computer-assisted instructional system the unmotivated student found scores and comments essential in working through programs, while the highly motivated students found these items much less important and concentrated more on speed and convenience. Thus, it becomes obvious that feedback cannot be categorically classified as being reinforcing. Instead, its reinforcing qualities appear to be a function of the characteristics of the particular student in relation to the programmed content. Further, a study by Krumboltz and Kiesler (1965) suggests that reinforcement is present in programmed instruction with or without feedback.

An additional problem with the reinforcing qualities of feedback is that programmed instruction does not make allowances for the student's previous knowledge in the particular subject area; rather all students must begin at the same point and are requested to overtly respond even though they may be quite familiar with the material and find it extremely easy. While Skinner (1958) saw this situation as advantageous, compelling the student to respond may actually impede the learning process. If programmed learning does involve the shaping of behavior (Skinner, 1958), requiring the previously knowledgeable student to work through the program involves the relearning of the sequential steps rather than the relearning of the final behavior. Consequently, such redundancy does not present the same situation as repeated trials involving the same response. The analogous laboratory animal situation would be exemplified by having a pigeon which has been trained to turn clockwise circles relearn this response through the original process of selectively reinforcing small approximations to clockwise circles. Such a procedure would be inefficient both in terms of the experimenter's time and in terms of the pigeon's time and would probably interfere with the pigeon's complete circle turning; yet this activity is what we ask students to do when we have them go through auto-instructional material involving content with which they are already familiar.

Hence, the study is a test of the hypothesis that the programmed material must be broken into small parts, each of which elicits a response which is subsequently reinforced. The specific problem, then, is to determine if efficiency of learning is dependent upon frame size. The study is exploratory in nature; no specific alternative hypothesis is being directly tested.

Related Research

Frame size defined as the amount of material presented between responses has received little attention in the literature. The "step-size" studies have usually been concerned with the difficulty of giving a correct response at any point in the program (Coulson and Silberman, 1960, 1961; Shay, 1961; Smith and Moore, 1962; and Kapel, 1965) rather than with the length of the material between responses. However, "step-size" has had at least four different meanings: (a) the length of the program frame (i.e., frame size as defined in the present study); (b) the difficulty of giving a correct answer (size of step between frames); (c) the complexity or length of the response required of the student before he receives feedback (Evans, 1965). These different usages have resulted in some confusion concerning the results of size of step" as they use it. Much of the literature on this topic has not bearing on the present study, and will be omitted here.

One type of study which superficially resembles the present one is the comparison of short-framed Skinnerian programs (extrinsically sequenced) with the longer-framed branching programs (intrinsically sequenced). However, unless unusually strict controls are effected, the contributions of frame size cannot be separated out from the other variables such as mode of responding and method of sequencing. As Crowder points out, "Linear and instrinsic programming, while having some superficial similarities, are basically different in approach, intention and rationale" (1963, p. 254). Consequently, studies of this type will not be considered.

Another type of study which appears to be similar is the comparison of auto-instructional programs to conventional text materials. While in a sense these studies are a comparison of different frame lengths, there are also qualitative differences between auto-instructional programs and conventional material. Conventional materials require no overt responding and, thus, require no feedback or controlled reinforcement. Thus, these studies are not strictly a study of different frame size.

In one example of a study in which auto-instruction was compared to a conventional text, Briggs et al. (1962) used four types of material to teach a lesson on the structure and the function of the United Nations. Two forms of the material were conventional (one with review material and one without) and two forms were Skinner-type programs (one requiring the first letter of the response word only, and the other requiring the complete response word). The investigators found that the conventional groups were more efficient in terms of time than were the Skinnerian groups. On the criterion tests (multiple choice) the Skinnerian groups excelled when the correct answers were identical to the responses learned in the program, but the conventional groups excelled when the correct answers were different than the responses in the programs. Briggs and his associates concluded that, "if you employ a Skinner-type program, do not expect even a moderate level of learning of program content which is not elicited during training" (p. 92).

In a second study reported by Briggs et al. (1962), one group of students worked through linear programs which required overt responding, while another group read the same material with the responses filled in. The reading group surpassed the overt group in time-efficiency for all of the materials, and surpassed them in test scores for the easy frames (based on the cues). This study suggests that for those frames which provide easy prompts for a student, overt responding is an ineffectual method compared to reading. In the present study, no attempt has been made to control prompting, but it was anticipated that the effect reported by Briggs et al. will be present across material versions representing different frame sizes.

Krumboltz and Kiesler (1965) studied the effects of partial reinforcement in programmed instruction by varying frame size as defined in the present study and by varying the availability of feedback. They used six versions of a rather short (177 - frames) program on interpreting educational test results. These versions were:

1. The standard program in which a question was asked on every frame with the answers provided.
2. A question on every frame with the answer provided for every fifth frame.
3. A question on every frame with the answer provided for every tenth frame.
4. A question only on every fifth frame with the answers provided.
5. A question only every tenth frame with the answers provided.
6. No questions asked with all frames consisting of declarative statements.

These versions presumably are arranged in descending order of the reinforcement which they provide. Criterion test results of the high school students in the study tended to support the hypothesis that the more reinforcement the better the test scores, with the exception that the mean for the fifth group was lower than that for the third, fourth, and sixth groups. The researchers did an analysis of variance on the test scores which produced a significant F ($F \doteq 3.05$; $df = 5, 77$; $p < .05$) for the test given immediately upon completion of the materials to half of the subjects. They did not, however, make individual comparisons of the means to determine which values differed. A second analysis of variance using scores on a delayed test (approximately two months) given to the other half of the subjects failed to produce a significant F.

Unfortunately, examination of the immediate test data presented by Krumboltz and Kiesler shows that the variances are not homogeneous. By squaring the reported standard deviations, computation of Hartley's F-max statistics gives a value of 9.46, which for six groups and 20 df, is significant at the .01 level.[1] While the analysis of variance F-test is usually robust in the face of a lack of homogeneity of variance, this robustness fails in the face of unequal sample sizes. Generally, a lack of homogeneity will inflate the F-value and produce significance where otherwise it would not exist (Cochran and Cox, 1957; Edwards, 1960). Hence, it is quite possible that if the scores were corrected for the lack of homogeneity, the obtained F would not be significant.

While the reported evidence suggests that the larger the frame size, the better the performance, this may be due only to poor statistical handling of the data. In addition, the experimenters had no control over the student's acquisition experiences and as reported above, the program was quite short. These considerations cast doubt on the validity of the findings.

In summary, there has been little research on the variable of frame size and the evidence is inconclusive. While Briggs (1968) in his review of "Learner Variables and Educational Media," states that the Skinnerian theory that small steps are needed has not gone unchallenged, he does not cite one study which is concerned with the length of the frame as defined in the present study. In 1962, Silberman stated that the research on programmed instruction at that time had barely scratched the surface and that "beyond demonstrating that a carefully written set of materials will teach if a student will spend enough time studying them, we have little unequivocal evidence for the principles of programmed instruction."

---

[1] Use of the F-max statistic assumes equal n's. Since the n's are unequal the largest number of degrees of freedom was used which gives a more powerful test than any other number. However, even using the mean number of degrees of freedom, 13.8, the F-max statistic is significant at the .01 level.

Although six years have passed since then, the situation has changed little. Many of the difficulties have been in the studies themselves; ample examples of these deficits are presented elsewhere (Silberman, 1962; Tuel, 1966; Flynn, 1967). Frequently, studies have not been well designed or well conducted and many extraneous sources of variation have confounded the results. A frequent weakness is that many investigators use extremely short programs which may not permit a definite pattern to establish itself or for novelty effects to dissipate.

Objectives

The purpose of the study was to explore an aspect of the general problem of identifying optimal learning materials for given individuals. The specific question being asked is the following: "Is there an optimal frame size for learning to occur in linearly sequenced auto-instructional programs?" Following from this question, "If the frame size in linearly programmed instructional material is varied while all other material-centered variables are held constant, will there be a differential efficiency of learning?."

CHAPTER II: PROCEDURES

Sample

Eighty-four students were randomly selected from the 10th grade and 84 from the 11th grade at Nova High School (Ft. Lauderdale, Florida).[1] Forty-two of the students in each grade are male and 42 are female. No attempt was made to representatively sample the student body according to any a priori selection of variables. However, each home room (AdCom room) was sampled in proportion to its relative size. According to school board policy, parents were notified by mail of the project and consequently had an opportunity to withdraw their children.

Nova High School is an experimental school in Broward County, Florida. The students are predominately from middle class homes in the Ft. Lauderdale area. This school was selected primarily because the independent study programs and modular scheduling permitted the needed flexibility to have the students participate in the study.

It was found at the time of the study that one tenth grade-girl who had been selected was not actually enrolled at the high school. In addition, a boy who was reported by his mother to be a non-reader was excused from participating. Of the remaining 166 students only 98 completed all parts of the study. The reasons for the attrition are presented in Appendix A. The resultant composition of the sample and the effects of the high attrition rate on the data are analyzed and discussed later in this chapter.

---

[1] The author is indebted to Dr. James E. Smith and Mr.David Fitzpatrick for allowing the students to participate in the study, and to the entire staff of Nova High School for their excellent cooperation.

Materials

A commercially prepared auto-instructional program on objective psychology was selected for use in the study (Holland and Skinner, 1961).[2] The criteria used in selecting a program were (1) that the subject be one with which high school students have a high probability of being unfamiliar; (2) that the length be such as to provide at least ten hours of instruction; and (3) that the program be representative of the principles of the Skinnerian approach (i.e., small frames, overt responses, and linear sequencing). The last criterion was included because only programs of this type can be readily adapted to the needs of the study. For example, if the program was branching, it would be extremely difficult to combine frames to make new frames. Further, branching programs generally have longer frames and multiple-choice responses.

Nine hundred sixty frames were selected from the first 24 sets of the program for use in the study. Ninty-six of these frames were used as introductory material to familiarize the students with the mode of instruction and were not used for collecting data. The remaining 864 frames were used for the study and provided the informational base against which data were collected. The introductory material was presented in a booklet separate from the other material. These booklets were designated Part I and Part II respectively. In addition to frames, Part I contains one "Exhibit" (i.e., graphs, tables, pictures, explanatory materials, etc.) and Part II contains 13 "Exhibits."

---

[2] The program was used with the permission of the publisher, McGraw-Hill, Inc.

The program was rewritten into seven versions with all variables except frame size per response held constant. These versions are:

(1) Small frames as originally written.

(2) Material of two of the smaller frames combined.

(3) Material of four of the smaller frames combined.

(4) Material of eight of the smaller frames combined.

(5) Material of 16 of the smaller frames combined.

(6) Material of 32 of the smaller frames combined.

(7) Material of all of the smaller frames combined so that no responding is required.

For illustration, the following are two adjacent original frames:

> A light flashed into the eye elicts constriction of the pupil. This sequence is called the pupillary________.
>
> In the pupillary reflex, a flash of light is said to _________the response.

These two frames are retained verbatim in version 1 of the material. In version 2 they would be modified as follows:

> A light flashed into the eye elicts constriction of the pupil. This sequence is called the pupillary reflex.
>
> In the pupillary reflex, a flash of light is said to_________the response.

The only change is that the response "reflex" is filled in for the first original frame. Consequently, only one response is now required for the two frames combined. The original spacing and format is maintained so that the only change is the frame size or frequency of responding. In the subsequent versions, responses only occur for the last original frame in each of the composite frames. Thus, each successive version, with the exception of version 7, requires only half as many responses as the preceding version. Version 7 requires no responses at all and is, therefore, qualitatively different than the other versions.

Examples of each of the seven versions are presented in Appendix B. All of the versions contain an equal number of the original frames (i.e., the same amount of instructional material).

Before the preparation of the different versions, the material was edited so that all the frames required only a single response instead of multiple responses as many did originally. The booklets were then prepared by arranging eight of the original frames on a page. Response blanks were filled-in with the correct answers or left blank according to the version as outlined above. Several frames were arbitrarily eliminated so that the total was evenly divisible by the 32 frames in version 6. In one case, a non-technical word which would probably have been unfamiliar to the students in the sample was changed to a more familiar word. Several of the original frames in the commercial program required the student to select a multiple-choice answer. These frames were rewritten to require a structured response by the student.

Design

The sample of 168 students was randomly divided into seven groups of 24 students each, with sex equally divided in each group. Each group worked through one of the versions of the material as outlined above under supervised classroom conditions. The students were unable to remove the materials from the classroom or to use them except at the designated time each day. Prior to beginning on the programs, the separate introductory material (Part I) for that particular version of the material was used to familiarize the student with that version. No data from the familiarization period were collected.

In order to minimize and counterbalance extraneous sources of variation, the original 168 students were physically divided into six groups of 28 students each. Four students within each of these physical groupings working with each of the seven material versions, so that the students in each version were distributed equally across the physical groupings. Each of the six groups was scheduled to meet for 64 minutes a day for the duration of the project (about two and one-half weeks). Each of the groups met at different times each day, rotating systematically through six blocks of time. Thus, all groups were given the opportunity of working at varying times throughout the day; no group had to work continually at a disadvantageous time, and no student had to miss the same regular class on successive days.

Because the project was superimposed upon the regular school activities, students were given the option of coming at different times than those for which they were scheduled. Students frequently selected this option,which resulted in further scrambling of the experimental groups. The result was that disturbances such as an electrical power failure which did occur during the study did not adversely affect any one experimental group alone, but affected some membersof all the experimental groups.

It was not feasible to obtain a single monitor to supervise the students for the duration of the study. Instead, a group of female adult volunteers was employed as monitors. There were five regular volunteers who rotated days and five substitute volunteers. Initially the monitors worked in pairs, and then later alone. Each of the students was exposed to all of the regular monitors and most of them to one or more of the substitutes.

Data Sources

Four sources of data were collected for the study. These are:

(1) An achievement test was given to each student the day after he completed the materials to determine his knowledge of the material. The development of this test is described in Chapter III and a copy of it appears in Appendix C.

(2) Completion time was recorded for each student to the nearest five minutes. This is the time which the student had the material, and not necessarily the actual time he worked. Occasionally times were not entered; in those

cases the time the student was scheduled was used. In general, this will inflate the time figure.

(3) The number of response errors made by each student in working through the materials was recorded daily by the students themselves. While this method allows room for students to falsify the records, it has been used with apparent satisfaction by other investigators (e.g., Kapel, 1965).

(4) A brief questionnaire was completed by the students immediately upon completion of the materials to determine their previous experience with psychology and their reaction to the materials. The questionnaire is discussed in Chapter III and a copy of it appears in Appendix C.

(5) A followup questionnaire was sent to all of the students after completion of the study to obtain information about the reasons for attrition and about the "Students'" opinions on some aspects of the study. The questionnaire is presented in Appendix C.

Additionally, SCAT scores were available in the school files for most of the students as an independent variable. The SCAT was given statewide in the ninth grade as part of a test battery. In general, when the SCAT scores are not available for a student, it means that he has transferred into Florida schools from out of state after the time that the SCAT was given for his class. Since the present sample consisted of tenth and eleventh grade students, the

SCAT scores were one and two years old for these students. Consequently, the SCAT scores may not adequately represent a measure of the students present intellectual capabilities. This fact should be kept in mind in interpreting the results.

Unfortunately, the scores obtained had been converted to percentiles based upon Florida norms and the raw scores were not available. In order to eliminate the underlying rectangular distribution, the percentiles were converted to normalized T scores (mean of 50, standard deviation of 10). This conversion did not necessarily normalize the sample used in the study, but only normalized the underlying distribution of the Florida norms. These transformed scores were used for all analyses involving the SCAT.

Attrition

The high rate of attrition reported earlier in this chapter gives cause for concern about the representiveness of the students remaining in the study. Consequently, considerable effort has been expended to ascertain the reasons for the attrition rate and to determine what effect it may have had on the data. Questionnaires were sent to all students to find out why they did not finish (See Appendix A). Table 1 reports the completion status of the 166 students who were to have participated in the study (excluding the two students discussed previously). As can be seen from this table, a total of 15 students did not start the study at all. The reasons for this failure to participate which were given by those students who responded to the questionnaire are presented in Appendix A. Unfortunately many of the students also declined to respond to the questionnaire. When permission was

Table 1

Completion Status of 166 Students Who Were Selected for the Study[a]

Version	Did Not Start	Participated But Did Not Finish	Finished All Parts of Study	Total
1	5	8	10	23
2	3	9	12	24
3	2	6	16	24
4	1	9	14	24
5	1	6	16	23
6	1	11	12	24
7	2	4	18	24
Total	15	53	98	166

[a] 168 students were originally selected for the study. One girl was not enrolled in school and one boy was excused because he was an aphasic non-reader.

granted for the study by the school administration officials, it was understood that the students must have the option of not participating. Consequently, no effort was made to force these 15 students to participate, and if such an effort had been made, it could have been detrimental to the study if the students already held or formed negative attitudes about participating. There is no reason to believe that the failure of these 15 students to participate was not random in regards to the treatment groups. Since they did not participate at all, they did not know that different treatment groups existed, and, of course, did not know to which one they were assigned.

Of the 151 students who began the study, only 98 or almost 65% completed the study through taking the achievement test. Seventeen or approximately 11% of the students finished reading the material, but did not return to take the test.[3] Six students or about 4% quit participating after only one or two sessions. The remaining 21% (31 students) participated most or all of the time, but did not complete the study. These three groups of students total 53 and are combined in column 3 of Table 1. The reasons for not finishing that are given by the students among these 53 who responded to the followup questionnaire are listed in Appendix A.

The high rate of attrition was in large part attributable to the timing of the study. Because of local school system problems, the starting date of the study was postponed one week from the originally scheduled date. This shift in the starting date caused the third week of the project to coincide with the final week of the high school's grading period. Consequently, many students who had not yet completed the study had to withdraw to complete their regular course requirements.

A χ^2 was caluclated for the data in columns 3 and 4 of Table 1 to determine the randomness of attrition across versions. The students in column 2 who did not start the study were omitted in part because the expected frequencies in all of the cells in that column are below 5. Further, however, it is assumed that the failure of these students to participate is random across versions as discussed above. For this reason, the frequencies in the second column were not added to those in the third column to calculate the χ^2. The obtained χ^2 is 5.48 (df = 6) which is not significant ($p > .40$). Hence, it is assumed that the students not completing the study are numerically randomly distributed across versions.

---

[3] Frequently stated reasons for failure to return for the test are absence from school and interference with other school work. See Appendix A for other stated reasons.

Of the students who completed the study, 28 were tenth-grade boys, 29 were tenth-grade girls, 18 were eleventh-grade boys, and 23 were eleventh-grade girls. Table 2 presents these data and reports two non-significant χ^2's. A χ^2 was calculated with the expected frequencies based upon the marginal totals to determine if a relationship exists between the grade level and the sex of the students who finished.

Table 2

Students by Sex and Grade Level Completing the Study

	Grade		
Sex	Tenth	Eleventh	Total
Male	28	18	46
Female	29	23	52
Total	57	41	98

Using marginal totals: $\chi^2 = .24$, d.f. = 1, n.s.

Assuming equal frequencies: $\chi^2 = 3.17$, d.f. = 1, n.s.

The calculated value is .24 (df = 1) which is not significant($p > .80$). Since sex and grade level were originally equally represented, a second χ^2 was calculated using equal expected frequencies in each cell (24.5) to determine if either grade or sex was unequally affected by the attrition. The value of this χ^2 is 3.17 and also is not significant ($p < .10$).

Similarly, Table 3 presents a contingency table showing the distribution of sex across versions of the students who completed the study. A χ^2 was calculated using the expected frequencies which are

Table 3

Students by Sex and Version Completing the Study

	Sex		
Version	Male	Female	Total
1	4	6	10
2	5	7	12
3	9	7	16
4	7	7	14
5	9	7	16
6	3	9	12
7	9	9	18
Total	46	52	98

Using marginal totals: $\chi^2 = 3.84$; d.f. = 6, n.s.

Assuming equal expectancies: $\chi^2 = 7.14$; d.f. = 6, n.s.

based upon the marginal totals to determine if the resultant frequencies of each sex are distributed evenly across versions. The value is 3.84 (df = 6) which is not significant ($p > .60$). A second χ^2 was calculated to ascertain if the resultant frequencies represent more than a chance departure from the originally intended equal cell frequencies. An expected frequency of seven was used for all cells and the value of χ^2 is 7.14 (df = 6). This value is not significant either ($p > .30$). Consequently, it appears that the attrition rate is reasonably uniform across the categories of sex, grade level, and treatment group and that it does not represent any trends which will adversely influence the data.

Because one of the problems contributing to the attrition was an apparent lack of time in the third week of the study, many of the slower reading students may have been forced out of the study. Also, students of lower ability would be more likely to need to direct all of their energies to their regular courses than would students of high ability (keeping in mind of course that the correlation between ability and high school achievement is far from perfect). Since students of low ability are frequently slow readers it would seem likely that these students would be still working on the materials in the third week and would have to drop from the study when pressed for time. Another factor influencing the speed with which the students completed the material is that each version requires a different number of responses. Students assigned to versions requiring a high frequency of overt responding should tend to take more time to complete the material than students who are assigned to materials with a low frequency of overt responding, because the responding itself will take more time.[4] Thus, it would seem likely the high frequency of response groups would suffer most from the attrition problem of the third week of the study. If this situation did occur the resulting data could no longer be considered random; rather an interaction would be present between ability (and/or working speed) and the material versions. Fortunately, the evidence does not indicate that this happened.

A χ^2 was calculated for the marginal frequencies for versions in Table 3 and a value of 3.43 (df = 6) was obtained. This χ^2 is not significant ($p > .70$), which indicates that the final frequencies for the versions are probably randomly distributed. Hence, it is unlikely that a significantly greater number of students was forced out of one version than was forced out of another by the third-week time squeeze.

Even though the numbers of students may be random across versions, the characteristics of the students such as ability or other learner-centered variables could be disproportionately distributed among the versions. Consequently, in order to better assess the students who completed the study, an analysis of variance was carried out for total SCAT scores in a participation by version interaction design. Two levels of participation--those who participated but did not complete the materials and those who did complete the materials--were used together with the seven versions. The 15 students who did not participate at all were excluded because it is assumed (as discussed above) that their nonparticipation was random across versions. It was reasoned that if the interaction is significant or if the versions' main effect is significant, then the attrition (or the original selection or both) has caused ability as measured by SCAT scores to be unequally distributed across the cells.

---

[4] The results of the analysis of the completion time data are presented in Chapter IV. The actual relationship between versions and completion time appears to be curvilinear across the range of versions studied; high response frequency requires high completion time, a minimum is reached around version 5, and then the times increase again.

SCAT scores were not available for all of the students as discussed earlier in this chapter. The cell sizes representing the available scores are listed in Table 4 together with the means and standard deviations. The summary of the analysis is presented in Table 5. Only the main effect for participation is significant, which implies that the main result of the attrition in regards to ability was to cut off the bottom of the distribution relatively evenly across versions.[5]

---

[5] The variances of the SCAT scores are not homogeneous. Hartley's maximum F-ratio test was used, and since the cell sizes vary the table was entered with the maximum degrees of freedom which should inflate the alpha level increasing the power of the test. The F-max value is 31.9 ($k = 14$, $df = 13$) which is significant at the .01 level. While the lack of homogeneity would not greatly influence the F test if the cell sizes were equal, the cell sizes are not equal. In general, a lack of homogeneity will inflate the value of F and will produce more significant results than would otherwise be obtained (Cochran and Cox, 1957 ; Edwards, 1960). Since the percentile scores are more homogeneous than the normalized scores (F-max value for percentile scores is 29.0), an analysis of variance using percentile scores was completed. This produced a F value of 11.5 ($p < .05$) for participation main effects and reduced F's for versions and the interaction. A square root transformation of the percentile scores ($\chi' = \sqrt{\chi}$) further reduced the lack of homogeneity (F-max = 20.3), but still did not eliminate it. An analysis of variance using the square root transformed scores produced an F value of 6.72 for paticipation main effects which is significant at the 5% level. The interaction and the main effect for versions were not significant and all the F values using the transformed scores are lower than those obtained with the original scores. If a transformation was found which made the variances homogeneous, the expected effect on the analysis

Table 4

Means and Standard Deviations[a] of the Available SCAT Scores (Total) of the Students Who Participated

Version		Not Completed	Completed	Total
	N	8	8	16
1	MEAN	52.7	56.2	54.5
	SD	12.0	8.5	10.7
2	N	7	10	17
	MEAN	51.4	61.3	57.2
	SD	6.1	7.5	8.5
	N	6	12	18
3	MEAN	56.0	57.2	56.8
	SD	7.1	7.3	7.2
4	N	8	14	22
	MEAN	51.0	61.0	57.4
	SD	10.4	12.6	12.8
	N	5	12	17
5	MEAN	48.4	57.6	54.9
	SD	10.2	8.0	9.7
	N	10	10	20
6	MEAN	50.9	58.0	54.4
	SD	10.9	7.7	10.1
	N	3	14	17
7	MEAN	58.0	57.1	57.2
	SD	2.2	8.3	7.5
	N	47	80	127
Total	MEAN	52.1	58.4	56.1
	SD	9.9	8.9	10.4

---

[a] Standard deviations reported in this table describe the sample rather than the population (i.e., the sums of squares were divided by n rather than n-1).

---

would be to further reduce the value of the F's. Since the principal reason for the analysis was to ascertain that the versions and interaction effects are not significant, and since this has been done, no further attempts were made to transform the scores.

Table 5

Analysis of Variance Summary Table:
Participation by Versions for SCAT Total Scores[a]

Source	DF	Sums of Squares	Mean Square	F
A (Participation)	1	1578.7	1578.7	13.1**
B (Version)	6	606.3	101.0	.8
AB	6	643.8	107.3	.9
error	113	13632.1	120.6	

[a] An unweighted-means analysis was used (Winer, 1962).

While this is an unfortunate occurrence, it is not as serious as if ability had interacted with the versions or if one or two versions had significantly different levels of ability from the other group. The tendency of losing the lower ability students from the study will require a redefinition of the population, and will perhaps limit the variance due to ability. Further, the distributions of SCAT scores may be positively skewed because of the loss of more low ability students than high ability ones.

Attrition may have also influenced the distribution across versions of students having different previous experiences with psychology. While no pre-test was given, the psychology questionnaire contained ten questions regarding the student's previous experiences with psychologists and with psychology. While all of these questions may not have the same intrinsic weight or importance, it was assumed that they did for the purpose of the following analysis. The number of questions marked "yes" by each student was used in a single classification analysis of variance to determine if the versions differed in the students' backgrounds. The means and standard deviations are reported in Table 6 (F-max = 3.8, K = 7, df = 17; n.s. at p = .05) and the summary of the analysis is shown in Table 7. As can be seen from the F value, there was a tendency for more variation to be present within the versions than between them.

Table 6

Means and Standard Deviations of Affirmative
Answers to Ten Psychology Background Questions

Versions	N	Means	Standard Deviations[a]
1	9	2.1	1.3
2	12	2.1	1.3
3	16	2.8	1.4
4	14	2.4	1.8
5	16	2.5	1.7
6	11	2.2	1.3
7	18	2.3	1.3
Total	96	3.4	1.4

[a] Standard deviations reported in this table describe the sample rather than the population (i.e., the sum of squares were divided by n rather than n-1).

Table 7

Analysis of Variance Summary Table:
Frequency of Affirmative Answers to Psychology Background Questions[a]

Source	DF	Sums of Squares	Mean Squares	F
Between	6	5.5	.92	.41 n.s.
Within	89	200.7	2.26	
Total	95	206.2		

[a] Questionnaires are not available for 2 students, so the analysis is based on the data of 96 respondents.

In general, individual questions received affirmative answers too infrequently to be analyzed statistically. For example, the question; "Have you ever been taught in school anything about psychology (besides the materials which you have just completed)?" received a total of 5 affirmative answers representing 3 versions.

Although the attrition rate was unduly high, there is no evidence that it has resulted in unrepresentative treatment groups, except for making them more homogeneous in regards to SCAT scores. Thus, while a much lower rate of attrition would have been desirable, there are enough students who have completed the study to provide meaningful information about the programmed materials.

Unfortunately tangible data are not available on student personality factors nor on their interest in the project. Consequently, no comparisons can be made between the students who dropped out and those who completed the study on affective and personality variables. Therefore, the possibility does exist that the treatment groups are biased on these dimensions. If interest in psychology was greater in some groups than in others it would completely distort the results of the study.

While direct evidence is not available, responses to two questions on the followup questionnaire (see Chapter III) give some indication that the dropouts do not differ appreciably from those who completed. The first question is, "If you were requested to participate in a similar study sometime in the future, would you?" Twenty-one students who participated but did not finish responded to this question (out of a possible 53) and 63 students who finished responded (out of a possible 98). Of the students who did not finish, 9 responded with "Yes" and 12 responded with either "No" or "Don't Know" (these categories were combined because of the small number of "No" responses--3). Of those who finished, 33 responded "Yes" and 30 responded "No" or "Don't Know." Calculation of χ^2 gives a value of .57 (df = 1) which is not significant ($p > .40$).

The second question is, "Do you think that a reward should be given to those who participate?" Of the students who did not complete the study, 16 answered "Yes" and 6 answered "No" while of those who did finish, 44 answered "Yes" and 19 answered "No." The χ^2 for these frequencies is .07 (df = 1) which is also not significant ($p > .70$).

If the students who finished differ substantially from those who did not, it would be expected that their attitudes toward future participation in research studies would also differ. Similarly, it would be expected that those who were least satisfied with the study would be those who would want a "reward." If that is the case, then satisfaction is not significantly different between the dropouts and those who finished. However, the situation is complicated by the fact that questionnaires are not available for all of either group and that those which were returned cannot be considered representative of those which were not returned. If not returning questionnaires is indicative of a negative attitude toward the study, then there is significant evidence that those who finished do differ in attitude from those who did not. Twenty-three out of 53 (43%) who participated but did not finish returned questionnaires while 63 out of 98 (64%) who finished returned questionnaires. χ^2 for returning questionnaires based upon these frequencies is significant ($\chi^2 = 6.15$; df = 1; $p < .05$), indicating that proportionately fewer persons returned the followup questionnaire among those who did not finish than among those who did. However, since the questionnaire was sent following the study, the different rate of questionnaire returns could be due to either preexisting attitudes or it could be a direct result of some aspect of the study. The available information is not sufficient to make this discrimination.

The most important concern with attitudes is whether or not they differ from group to group. The data for the above questions are presented in the analysis section (Chapter IV); expected cell frequencies are too small to allow calculation of χ^2's, but no trends are evident. While there is variation among the treatment groups, the variation is different for different questions. The numbers returning questionnaires in each group are listed in Chapter IV. A χ^2 for those returning questionnaires versus those not returning them across versions gives a value of 8.0 ($df = 6$; $p > .20$). While this non-significant value indicates that the return of questionnaires is fairly even across versions, it is confounded by the fact that 13 returned questionnaires could not be identified according to treatment group and are excluded from the χ^2. These could be distributed proportionately across version or could be heavily concentrated in a few versions. Also, the problem still exists that the reasons for not returning the questionnaire can be a direct result of the study or can be unrelated to the study. Consequently, little can be concluded about attitudes.

CHAPTER III: INSTRUMENT DEVELOPMENT

Three instruments were developed for use in the present study. One of these was a 64-item achievement test, the second a 14-item "psychological questionnaire," and the third a questionnaire containing items concerning the attrition rate and overall attitudes toward the project. The development of these instruments is described below and the final versions are in Appendix C.

An achievement test was developed to measure the effectiveness of the instructional material. The test is identical for each version of the material and is in a multiple-choice format. The test covers only Part II of the materials and not Part I. Rather than prepare a complete table of specifications for the material, the nature of the programmed instructional material being in small frames permitted a random sampling of material to be tested. This method was certainly more efficient than preparing a table of specifications and it probably more validly represents the material. In order to equally represent each version of the material, 108 of the original frames were randomly selected from Part II of the material - one from each page. This represents 1/8 of the total number of frames. Then a multiple-choice item was written to reflect the content of each of the selected frames.

The completed questions were then submitted to two judges who critically read them to detect ambiguities, inaccurate statements, and poor distracters. Where possible, the judges suggested methods of improving weak questions.

Since the test had to be given within a sixty-four-minute period, only 64 questions are included in the final version. The 64 questions represent 7.4% of the total possible frames in Part II. These questions are allocated in such a way that they represent also 7.4% of the response frames for each version of the material in which responding is required. Table 8 illustrates the numerical allocation of these questions.

Table 8

Frequency of Test Frames Corresponding to Response Frames in Each Version

			Test Frames Coinciding with Response Frames Before Item Elimination		Test Frames Coinciding with Response Frames After Item Elimination	
Version	No. of Frames per Response	Total No. of Response Frames	Number	Percentage	Number	Percentage
1	1	864	64	7.4	55	6.4
2	2	432	32	7.4	28	6.5
3	4	216	16	7.4	13	6.0
4	8	108	8	7.4	6	5.6
5	16	54	4	7.4	3	5.6
6	32	27	2	7.4	1	3.7
7	NO RESPONDING REQUIRED		———	———	———	———

Additionally, each of the eight paginal frame positions is equally represented by eight test items.

The selection of the questions in the final version was based upon the following criteria:

1. Each of the responding versions must be equally represented as described above.
2. Poor quality questions as indicated by the judges were not used.
3. Questions were physically distributed as much as possible across the material for each of the responding frames.

Where these criteria were not adequate to choose between two or more questions, the choice was made randomly.

The resultant 64 questions were then randomly ordered for the final version to eliminate the existence of any temporal ordering effect in terms of the amount of time which had passed since that part of material was studied. The students recorded their answers to the questions on mark-sense cards which were electrically scored.

After administration of the test, the questions were subjected to an item analysis so that weak and nondiscriminatory items could be eliminated. A difficulty index was calculated for each item using the scores of all of the students. Then a discrimination index was calculated for each item by subtracting the number of correct responses in the lower 27% of the students from the number of correct responses in the upper 27% of the students, and then dividing this difference by the maximum possible difference.[1] These two indices are reported for each item in Table 9.

---

[1] While there are many item analysis techniques which are more so-

Table 9

Item Analysis of the Psychology Achievement Test[a]

Item	Difficulty[b] Index	Discrimination[c] Index	Item	Difficulty[b] Index	Discrimination[c] Index
1	57	.44	33	76	.33
2	85	.41	34	42	.30
3	88	.33	35	77	.44
4	64	.59	36	61	.61
5	69	.48	37	74	.48
6	67	.67	38	36	.41
7	25	.33	39	64	.63
8	44	.74	40	58	.63
*9	92	.18	41	63	.33
10	57	.59	*42	10	0
11	42	.52	43	65	.52
12	82	.44	44	68	.48
13	52	.59	45	82	.33
14	80	.48	*46	94	.07
15	86	.37	47	74	.70
16	28	.59	48	67	.74
17	71	.52	49	54	.74
18	75	.44	50	40	.59
19	77	.30	51	77	.67
20	61	.59	52	71	.41
21	54	.78	53	68	.67
22	80	.30	54	68	.59
*23	88	.18	*55	65	.26
24	82	.33	56	59	.67
25	71	.37	*57	97	.15
*26	89	.26	*58	93	.22
27	40	.48	59	61	.41
28	54	.85	60	79	.52
29	56	.78	61	79	.44
30	35	.37	62	56	.70
31	58	.56	63	36	.41
32	63	.63	*64	93	.22

[a] Based on scores of 98 students; standard deviation (before item elimination) is 11.7.

[b] Difficulty index is percentage of all students getting item correct.

[c] Discrimination index is found by $\frac{U-L}{N}$ where U is the number of students in the upper 27% getting the item correct, L is the number of students in lower 27% getting item correct, and N is 27% of the students.

*Indicates items with a discrimination index less than .30. These items were eliminated from the final test version.

A discrimination index of .30 or above was selected as the cutoff criterion for the retention of test items.[2]

The difficulty index was not used to eliminate items because the extremely easy and the extremely difficult items contribute very little to determining the composition of the top and bottom 27%. Further, the discrimination index removes the items of high and low difficulty as well as the other items which do not discriminate. Using the discrimination index criterion, nine items were eliminated. Each of these items is marked with an asterick in Table 9. After removal of the nine items, the membership of the top and bottom 27% changed by only one student, which indicates that there is no need to recalculate the discrimination index based on the revised scores.

The effect of eliminating these items on the representative sampling of frames is shown in Table 8. Student scores were then calculated on the basis of the remaining 55 items and used in the analysis reported in Chapter IV.

---

phisticated than the discrimination index, there is evidence that the use of a discrimination index is quite effective. Engelhart (1965) empirically compared discrimination indices based on the upper and lower thirds and on a median split with tetrachoric, phi, biserial and point-biserial correlations and with the Davis method. The discrimination indices were extremely effective and gave a better indication of the actual number of discriminations than did the correlational techniques.

[2] Ebel (1965) recommends that items with indices of discrimination between .20 and .29 are "marginal items, usually needing and being subject to improvement," and that items with indices below .19 are "poor items to be rejected or improved by revision." Five items fall in the former category and four in the latter.

Since the test was administered only once, test-retest reliability could not be calculated. Instead, internal consistency was determined using Kuder-Richardson formula 20, as follows:

$$r = \frac{n}{n-1} \; \frac{S_t2 \qquad \Sigma\, p_iq_i}{S_t2}$$

where:

n is the number of items,

S_t2 is the variance of the test scores,

p_i is the proportion of correct responses on Item i

and q_i is the proportion of incorrect responses on Item i.

Using this formula, the reliability was found to be .94. This coefficient is considered to be more than adequate for the test to be considered reliable.

No attempt was made to determine criterion-related validity of the instrument, since it is an achievement test with high apparent content validity. Each question was written to reflect the content of a specific randomly selected frame. Since the internal consistency was relatively high, it is reasonable to assume that the test tends to uniformly measure the content of the programmed materials, which implies high content validity.

The second instrument used in the study was a "psychology questionnaire," which was designed to serve two purposes. Since a pretest was not given prior to the study in order to avoid practice effects and to avoid establishing a set for the material to follow, a crude attempt was made in this questionnaire to determine if the students had any prior acquaintance with psychology. Thus, 10 of the questions on the student questionnaire reflect this end. The remaining four questions attempt to assess the students' reaction to the material. No validity or reliability data are available for any of these questions and they should be regarded only as an indication of the students' reactions. Hence, they should be interpreted cautiously. The items were worded in question form and the students responded by circling "Yes" or "No" to each question.

The rationale for using "Yes-No" questions (i.e., binary choice) rather than some other type as 5-point Likert scales, is based in part upon the nature of some of the questions. For example, the question, "Is your father or mother a psychologist?" presents only binary alternatives. Other items, however, could be used with a larger number of rating points. If the questionnaire was composed of these two types of items, it would be difficult if not impossible to create a meaningful composite of the items. There is no justification for summing scales with differing numbers of points.

For the items dealing with attitudes toward the material, scales utilizing more than two points could be used. The variance would be increased and finer discriminations could be made among individuals than with the two point scale. However, the Broward County School Board has adopted a strong position against psychological and attitudinal testing. Part of the testing policy adopted in June, 1967 reads as follows: "No tests other than standardized intelligence, aptitude and achievement tests shall be scheduled or administered without the express consent of the Board." The policy also provides that if a test is approved, parents must be notified by mail on School Board letterhead of the testing and must be given the opportunity to inspect the instrument prior to the testing. Further, it gives the parents the right to remove their children from the testing. Faced with these restrictions, it was decided that a simple "Yes-No" Questionnaire could be administered without objections, whereas a sophisticated-looking rating scale could not. Thus, both the limited number of items and the binary response mode reflect the local ban on psychological tests.

These same considerations are also reflected in the third instrument which has a "Yes-No" format. The primary purpose of the third instrument was to ascertain the reasons for attrition. (The attrition questions and the results are listed in Appendix A.) The other questions are largely based on concerns raised in conversations with the students in the study and on indirect feedback from the students. These questionnaires were given to the students after the completion of the study and were returned to the investigator by mail. In an attempt to secure a good rate of return and to facilitate frankness, the students had the option of not putting their names on them. Consequently, some of the questionnaires cannot be identified and they are excluded from analyses where identification is necessary (e. g., comparisons across versions).

CHAPTER IV: DATA ANALYSIS

Although, there seems to be no difference among the final samples on the various independent variables (see Chapter II), the unexpectedly high attrition rate for the subjects nevertheless has limited the possible analyses and interpretations because of the reduced n's. For example, interaction designs which can be employed, must be limited to two way designs. Also, of course, the low n's reduce the stability of the estimated parameters used in the analyses. Further, since the resultant cell sizes are unequal, homogeneity of variance and normal error distribution cannot be assumed.

Criterion Test Scores

Table 10 presents the resultant sample sizes by material versions and by sex along with the means and standard deviations for the psychology achievement test. As can be seen, the n's vary from a low of 3 to a maximum of 9. An inspection of the standard deviations especially when broken down by sex -- strongly suggests that the samples do not have homogeneous variances. If the sample sizes were equal, a lack of homogeneity would not seriously affect the F-test in an analysis of variance. However, since the sample sizes do vary greatly, Bartlett's test of homogeneity was used. The calculated value of χ^2 by this test is 7.05 which is not significant (d.f. = 13, $p > .80$). This indicates that homogeneity of the

Table 10

Means and Standard Deviations[a] of the Psychology Achievement Test by Version and Sex

Version		Male	Female	Total
1	N	4	6	10
	MEAN	41.0	35.3	37.6
	SD	12.0	13.0	12.9
2	N	5	7	12
	MEAN	35.8	38.9	37.6
	SD	8.8	9.4	9.3
3	N	9	7	16
	MEAN	30.3	34.0	31.9
	SD	9.3	10.3	9.9
4	N	7	7	14
	MEAN	32.7	33.4	33.1
	SD	7.8	14.0	11.3
5	N	9	7	16
	MEAN	29.2	35.0	31.8
	SD	15.1	8.5	12.9
6	N	3	9	12
	MEAN	25.7	40.2	36.6
	SD	16.4	8.2	12.5
7	N	9	9	18
	MEAN	31.7	35.8	33.7
	SD	8.9	12.3	10.9
Total	N	46	52	98
	MEAN	32.0	36.2	34.2
	SD	11.7	11.2	11.6

[a] Standard deviations reported in this table describe the sample rather than the population (i.e., the sums of squares were divided by n rather than n-1).

variances is probably present.

A second problem, also due to the attrition rate, is that normality cannot necessarily be assumed for the samples. The F-test is not greatly affected by non-normal distributions when the sample sizes are relatively large, but unfortunately, the sample sizes are rather small. However, tests of homogeneity of variance such as the Bartlett test used are usually extremely sensitive to departures from normality (Hays, 1963). Hence, since Bartlett's test produced a non-significant χ^2 it is reasonable to assume that the distributions approximate normality and that they will not greatly affect the F-test. Still, in drawing conclusions from the data, it must be recognized that samples with small n's, especially those of 3 or 4, cannot closely represent normal distributions.

Since the assumptions of the F-test appear to have been met, a two-way interaction (sex by version) analysis of variance was completed for the psychology achievement test scores. The summary table for this analysis is presented in Table 11. As can be seen from the table, neither the main effects nor the interaction are significant. In fact, for the versions' main effect and for the interaction term, the mean squares are smaller than the error term. This phenomenon indicates that the variance due to the differences among the individuals tended to be greater than the variance due to the effects of having different versions of the material. In other words, even with the variance due to sex removed, the seven versions of the material did not cause differential performance on the achievement test.

Table 11

Analysis of Variance Summary Table
Sex X Versions for Psychology Achievement Test Scores

Source	DF	Sums of Squares[a]	Mean Square	F
A (sex)	1	308.4	308.4	2.17 N.S
B (version)	6	467.9	78.0	.55 N.S.
AB	6	687.0	114.5	.81 N.S.
error	84	11933.3	142.1	

[a] An unweighted-means analysis was used (Winer, 1962).

A possible factor influencing the criterion test scores could be ability if it is differentially distributed across versions. Table 5 in Chapter II presents a summary of an interaction analysis of SCAT total scores, in which no difference was found for the versions' main effect or for the interaction of versions with the participation variable. Thus, it would be unlikely that the SCAT scores are differentially distributed across the versions.

Nevertheless, as a further check, an analysis of covariance was made using the SCAT total scores as the covariate. The Pearson correlation of the criterion test scores with the SCAT total test scores for the 80 students for whom SCAT scores were available is .58.[1] For an N of 80, this value is significant beyond the .001

---

[1] The correlation of the criterion test scores with the SCAT verbal scores is .62 and with the SCAT quantitative test scores is .35. Since the correlation coefficients of the verbal and of the total SCAT scores

level, indicating a reasonably strong relationship between the SCAT and the psychology achievement test scores (the coefficient of determination is .34).

The relationship suggests that an analysis of covariance using the SCAT scores as the covariate might be appropriate and might permit the effects of different versions to be detected after the effect of intelligence (as measured by the SCAT) has been statistically removed. The analysis of covariance, then, will statistically equalize the group on this variable if they do differ.

Although the factor of sex was not significant in the above analysis of variance and the point biserial correlation of sex with the SCAT total scores is .04, sex could still interact with intelligence to influence the achievement test scores. However, reference to the Table 12 will show that the resultant n for males in version 6 is 2, and that four other cells have n's of 4. Consequently, if the analysis of covariance was computed with sex as a factor in an interaction design, some of the regression lines would be based on extremely small and probably highly unreliable samples. Sex, then, is ignored as a factor, and the analysis of covariance is completed only for the versions in a

---

are essentially identical, the verbal scores could probably be used interchangeably with the total score in the analysis of covariance. The total was chosen to use because it can be regarded as a measure of general intelligence, correlating with the WAIS total at about .84 (Green, 1965).

The internal consistency reliability of the criterion test was reported in Chapter III to be .94 and the internal consistency of the SCAT is .95 (Green, 1965). Consequently, because of these high reliabilities, correcting the correlation coefficient of .58 for attenuation yields a coefficient of only .61.

Table 12

The Number of Students by Sex and Version for Whom Both the Achievement Test Scores and SCAT Scores Were Available.

	Sex		
Version	Male	Female	Total
1	4	4	8
2	4	6	10
3	8	4	12
4	7	7	14
5	7	5	12
6	2	8	10
7	7	7	14
Total	39	41	80

single classification design.

Table 13 gives the means and standard deviations of the SCAT total and of the criterion test for the samples used in the analysis of covariance. Hartley's maximum F - ratio was used to test for homogeneity of variance.[2] For the SCAT the ratio of the largest to the

---

[2] Hartley's test assumes that the degrees of freedom are equal for all cells which is not true for the data in this case. Consequently, in determining the critical value of F-max, the largest number of degrees of freedom, 13, was used to enter the table. This procedure should have the effect of increasing the size of alpha beyond the .05 level, and thus, increasing the likelihood of rejecting the null hypothesis of equal variances.

Table 13

Means and Standard Deviations[a] of the SCAT and the Achievement Test

Version	N	SCAT		Achievement	
		Mean	SD	Mean	SD
1	8	56.2	8.5	35.5	13.5
2	10	61.3	7.5	38.8	9.6
3	12	57.2	7.3	32.9	9.5
4	14	61.0	12.6	33.1	11.3
5	12	57.6	8.0	35.2	12.7
6	10	58.0	7.7	38.8	10.2
7	14	57.1	8.3	34.4	8.7
Total	80	58.4	8.9	35.3	10.8

[a] Standard deviations reported in this table describe the sample rather than the population (i.e., the sum of squares were divided by η rather than η-1).

smallest variance is 3.0 (K = 7, df = 13) which is not significant ($p > .05$). For the criterion test, the ratio is 2.4 (K = 7, df = 13) which also is not significant ($p > .05$). Consequently, it is assumed that the variances are homogeneous.[3]

A test was made for homogeneity of within sample regression and the calculated F is .97 (df = 6, 66), which, of course, is not significant. Consequently, it is assumed that the regression is homogeneous and the treatment means were tested for differences. The summary of the analysis of covariance is presented in Table 14. The calculated value of F is .91 which is not significant. Thus, it seems that the SCAT scores are equally distributed across versions, since removing their effects has no significant influence on the treatment means.

Table 14

Analysis of Covariance for the Psychology Achievement Test Scores with SCAT Scores as the Covariate.

Source	DF	Sums of Squares	Mean Square	F
Treatments	6	453.6	75.6	.91 N.S.
Error	72	5983.8	83.1	
Total	78	6437.4		

---

[3] Footnote 5 in Chapter II stated that the variances of the SCAT scores were not homogeneous. However, in that test the smallest and largest variances were both in the non-completion groups. These groups were, of course, not included in the analysis of covariance.

Earlier in this chapter, it was reported that Bartlett's test failed to detect heterogeneity of variance for the criterion test. A second test was performed here because only 80 subjects are used for the analysis of covariance instead of all 98.

Completion Time

Completion times to work through the programmed materials were recorded in minutes for each student to the nearest five minutes. These times, however, should be considered as a rough estimate because of errors in the recording. For example, one day the monitor failed to record the times for the majority of the students. These times were subsequently estimated from the students' own records and from the times the students were originally scheduled to work with the project.

The means and standard deviations for each of the versions are presented in Table 15. The F-max statistic was used again to test

Table 15

Means and Standard Deviations of Completion Times in Minutes

Version	N	Mean	Standard Deviation[a]
1	10	347.0	51.3
2	12	306.2	60.8
3	16	281.6	54.9
4	14	240.4	63.7
5	16	231.9	64.2
6	12	260.0	48.0
7	18	276.1	70.9
Total	98		

[a] Standard deviations reported in this table describe the sample rather than the population (i.e., the sum of squares were divided by n rather than n-1).

for homogeneity of variance. The calculated value is 1.9 (K = 7, df = 17) which is not significant (p > .05). An analysis of variance was then performed on the data and the results are reported in Table 16.

Table 16

Analysis of Variance for Completion Time

Source	DF	Sums of Squares	Mean Square	F
Between	6	105515.6	17585.9	4.31**
Within	91	371096.9	4077.9	
Total	97	476612.5		

** Indicates significance at p < .01

The F value of 4.31 (df = 6, 91) is highly significant (p < .01). This finding was anticipated because it was assumed that it would take longer to complete material which required a lot of overt responding than it would to complete with little or no overt responding. However, inspection of the distribution of means does not permit that simple an explanation. The means decrease from a high for version 1 to a low for version 5, and then increase again. In order to determine which differences among the means are statistically significant, a comparison among the means was made using the test

$$F = \frac{\left(\overline{X}_1 - \overline{X}_2\right)^2}{MS_{\text{error}} \left[1 + \frac{\Sigma X_t^2}{(k - 1)\ \Sigma X_w^2}\right] \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

where

X_1 and X_2 are the adjusted treatment means,

ΣX_t^2 is the total sum of squares for X and

ΣX_w^2 is the within sum of squares for X.[4]

Table 17

Differences Between Completion Time Means

Version		1	2	3	4	5	6	7
	Means N	10	12	16	14	16	12	18
1	347.0	-	40.8	65.4*	100.1***	108.9***	87.0**	70.9 *
2	306.2		-	24.6	59.3*	68.1*	46.2	30.1
3	281.6			-	34.7	43.5	21.6	5.5
4	246.9				-	8.8	13.1	29.2
5	238.1					-	21.9	38.0
6	260.0						-	16.1
7	276.1							-

* Indicates significance at $p < .05$

** Indicates significance at $p < .01$

*** Indicates significance at $p < .001$

The results of the comparisons are presented in Table 17. The differences between the mean for version 5 and those for version 6 and 7 are not significant. However, the finding of increasing completion time after version 5 still may be a true function of the versions rather than attributable to sampling error.

---

[4] This test statistic is discussed in Winer (1962), Cochran and Cox (1957), Snedecor (1956), and other books dealing with multiple group comparison.

Descriptively, the correlation ratio η will yield an index of the strength of the relationship regardless of the shape of that relationship. Using the information in Table 16, η is calculated by taking the square root of the ratio of the between sums of squares to the total sums of squares. This value is .47.

Pearson product-moment correlation measures the strength of the linear relationship between two variables. For the relationship of versions and completion time, the Pearson coefficient is -.30 (the negative sign is merely a function of the numbering of the versions). The test of the null hypothesis that $\eta = r$ gives an F value of 3.03 (df = 5, 91) which is significant at the .05 level[5] This implies that the relationship between versions and completion time is best described by a curvilinear function. Unfortunately, η's main importance is as a description of the sample and not as an estimate of the population association (Hays, 1963). The calculation of η has only confirmed the obvious fact that the completion time is not linearly related to versions.

The estimation of the population relationship and the determination of the best fitting function is usually done by trend analysis. However, trend analysis normally assumes equal intervals between the levels of the independent variables. Such as assumption cannot be made with the material versions because the frame size doubles for each successive version. Because the relationship of successive versions is known, it would seem reasonable that a corrective function could be developed a priori to equalize the intervals. Thus, if the logical assumption of a linear relationship prevailed, some equation such as

---

[5] $F = [(\eta^2 - r^2) / (n - 2)] / [(1 - \eta^2) / (N - n)]$

$$T = R + f(F)$$

could be utilized, where T is the completion time, R is the reading time and f(F) is a function of the number of response frames (F) in the material. The logical assumption would be that f(F) would actually be the time, t, required to make a response multiplied by the number of responses (i.e., tF). Then version 7 would have completion time, R, since no responding is required; version 6 would have completion time, R + 27t, since there are 27 elicited responses; version 5's completion time would be R + 54t and so on. (Since the test of $\eta = r$ indicates curvilinearity, the function, f(F) is probably not as simple as tF. Rather, some function such as tF^p where p is an unknown exponent would be more likely).

Unfortunately, this simple relationship does not work because the completion for versions 4, 5, and 6 are all less than that for version 7. Thus, if the model is true, the data imply

$$R > R + 27t > R + 54t > R + 108t$$

and

$$R < R + 216t < R + 432t < R + 864t$$

These contradictory relationships are, of course, not possible. Even without the contradictions, the first set of inequalities requires t, the time to respond, to be a negative quantity, which of course, is also impossible.

Consequently, as was suggested previously by η the simplest a priori relationship between completion time and versions does not fit the data, and a simple correction for the lack of equal intervals is not available. Whether the curved nature of the completion time is due to extraneous sources of variation such as sampling errors or to the effects of the frame sizes cannot be discerned from the present study. All that can be concluded from the comparisons reported in Table 17 is that the versions requiring a high frequency of responding (i.e., versions 1 and 2) do have significantly longer completion times than do versions requiring lower completion times.

Error rate

Each student kept a record of the numbers of errors he made in working through Part II of the material. Since the frequency of responding varied with the material version, the error rate is in part a function of the version. The raw error means for each version are reported in Table 18. Inspection of the table shows that the mean errors decrease as the number of responses decrease except for version 2 which increases slightly. In order to allow for the different maximum number of errors, the errors were converted to percentages of responses as shown in Column 5 of the table. Expressed as percentages, the means increase steadily across versions to a high for version 6.

In order to better understand the function of the errors, the correlations of both the raw scores and the percentage errors with the criterion test were calculated. Between raw score errors and the criterion test, the Pearson correlation coefficient is -.16, which is not significant (df = 96, $p > .05$). Between percentage errors

Table 18

Means and Standard Deviation[a] of Errors Made in the Programmed Materials

Version[b]	Number of Responses	Raw Error Mean	Raw Error SD	Percent Error Mean	Percent Error SD
1	864	36.4	22.5	4.2	1.1
2	432	36.9	33.9	8.3	7.9
3	216	21.4	18.8	9.9	8.9
4	108	11.5	8.5	10.6	7.8
5	54	8.1	4.8	15.0	8.9
6	27	4.2	4.5	15.6	16.7

[a] Standard deviations reported in this table describe the sample rather than the population (i.e., the sum of squares were divided by n rather than n-1.).

[b] Version 7 is omitted since no responses were required.

and the criterion test, the correlation is -.36, which is significant beyond the .001 level. Additionally, the errors were converted to T-scores (mean = 50, sd = 10) within each version, and the correlation with the criterion test was found to be -.32, which is also significant ($p < .01$). These correlations indicate, then, that a negative relationship exist between the errors made by a student and his score on the criterion test. However, even with the significant correlations, only 13% of the variance of the criterion test is in common with the percentage errors (only 10% with the T-scores).

The number of errors made by an individual should be in part a function of his mental ability. Thus, at least some of the relationship between errors and the criterion test may be due to intelligence.

The correlation between percentage of errors and the SCAT total scores is -.25. A correlation coefficient of .58 between SCAT total scores and the criterion test scores was reported earlier in this chapter. Removing the correlation with the SCAT, the partial correlation between errors (percentage) and the criterion test is - .27. While this value is still significant ($p < .01$) the variance held in common is reduced to about 7%. Considering the relatively small magnitude of the relationship between errors and the criterion test, it appears that error rate alone is a poor predictor of performance on the criterion test. The intercorrelation of the various variables is discussed further in the next section.

Intercorrelation of Dependent and Independent Variables

Intercorrelation coefficients for the criterion test, completion time, and percentage errors along with six independent variables are presented in Table 19. The correlations of the criterion test scores with other variables are particularly interesting. The lowest correlation (-.03) is with versions which gives further indication of the lack of a relationship between the test and the frame sizes. On the other hand, the criterion test has significant correlations with the three SCAT scores, with sex, and with percentage errors. These coefficients suggest, as discussed previously, that individual differences are the most important determiners of the test scores.

The correlation (-.36) with percentage errors is difficult to explain, since percentage errors correlate significantly (.34) with versions even though the criterion test and versions have essentially no correlation. If a multiple R was being calculated with the criterion

Table 19

Intercorrelations of the Criterion Variables
With Several Independent Variables[a]

	Criterion Variables			Independent Variables					
	1	2	3	4	5	6	7	8	9
1. Criterion Test		.11 (98)	-.36** (80)	.62** (80)	.35** (80)	.58** (80)	-.23* (98)	-.16 (98)	-.03 (98)
2. Completion Time (minutes)			-.01 (80)	-.21 (80)	-.24* (80)	-.21 (80)	-.02 (98)	-.12 (98)	-.30** (98)
3. % Errors[b]				-.28* (66)	-.12 (66)	-.25 (66)	.07 (80)	.19 (80)	.34** (80)
4. SCAT V					.62** (138)	.88** (138)	-.05 (138)	.08 (138)	.11 (138)
5. SCAT Q						.88** (140)	.15 (140)	.00 (140)	.13 (140)
6. SCAT T							.04 (140)	.02 (140)	.07 (140)
7. Sex								-.01 (166)	.03 (166)
8. Grade									.02 (166)
9. Version[c]									

[a] The n for each correlation is reported in ( ) below each coefficient

[b] Since no responding was elicited in version 7, correlations with % errors are based only on versions 1 through 6.

[c] Versions was quantified for the calculation of the correlation coefficients by utilizing the numbers 1 through 7 which have been used nominally in this study. These numbers are ordered in terms of frame size, but they do not reflect the differences in frame size. The total number of responses in the material or the number of original frames per response might be better indices of the versions variable.

* Indicates significance at $p < .05$

** Indicates significance at $p < .01$

test as the predicted variable, versions would act as a suppressor variable as long as percentage errors was in the equation (but the increase in R would be small because of the low magnitudes of the correlations). This relationship implies that the version influences the error rate, but that this influence is unrelated to the criterion test. When the variance in the error rate (percentage error) due to the effect of versions is partialled out, the relationship of error rate to the criterion test is increased. (It was pointed out previously that atleast some of the relationship between error rate and the criterion test is due to mental ability).

The intercorrelations involving the three SCAT scores, sex, grade, and versions reflect the original sampling before attrition. The coefficients for sex, grade and versions differ slightly from zero only because two of the original 168 students are not included and because of rounding errors. The correlations involving the SCAT scores and the other three variables are based upon the students for whom SCAT scores were available. These coefficients are surprisingly low--especially with versions--since the SCAT scores were not taken into consideration in sampling. The correlation of SCAT quantitative with sex is significant ($r = .15$, $p < .05$) and the correlation with versions approaches significance ($r = .13$, $p < .10$) but in both cases only about 2% of the variance is held in common. These figures, however, cannot be interpreted as an indication of the representiveness of the students who completed the study; they only have implications for the original sample.

Another correlation coefficient which requires comment is that between completion time and percentage error rate (-.01). Percentage error rate is reported in Table 19 instead of raw error rate because the raw error rate strongly reflects the different response frequencies which distinguish the different versions. However, use of the percentage error rate may cloud the true relationship between error rate and completion time. Completion time should be a function of the number of errors made since making errors should be time consuming (this factor was omitted in the previous discussion of completion time because error rate had not yet been discussed and because time consumed by errors still does not explain the curved nature of the time and versions relationship). The correlation of raw error rate with completion time is .39 which is, of course, significant. Part of this relationship can be explained away by the fact that both error rate and completion time are strongly influenced by the frame sizes. The correlation of raw error rate with versions (quantified by ranking only) is -.54 and the correlation of completion time with versions is -.30. Thus, the partial correlation of raw error with completion time is .29, which suggests that completion time is a function of error rate, in addition to reading time and number of responses.

Based upon this information, the a priori equation for completion time can now be modified to include error rate. Still assuming linearity, the equation becomes:

$$T = R + tF + eE$$

where e is the time required to make an error and E is the number of errors made. In order to assess the contribution of e, a system of three equations were set up using the mean completion times, mean error rates, and the number of responses for versions 1, 2, and 3. The resultant equations are:

$$\begin{aligned} 347.0 &= R + 864t + 36.4e \\ 306.2 &= R + 432t + 36.9e \\ 281.6 &= R + 216t + 21.4e \end{aligned}$$

Since completion time is essentially linear considering only versions 1,2, and 3, these equations were solved simultaneously giving the following values:

$$\begin{aligned} R &= 259.6 \text{ min.} \\ r &= .10 \text{ min.} \\ e &= .03 \text{ min.} \end{aligned}$$

While the curvilinear nature of the relationship across all versions tends to negate the premises these values are based upon, they should, nevertheless, give an indication of the relative magnitudes of the three time factors. As can be seen from the value for e, 33 + errors would have to be made before a minute is added to the completion time. Since the largest mean number of errors is 36.9, error rate does not seem to be a major factor in completion time. Further, since error rate is not independent of the number of response elicitations, there is little to be gained from having it in the equation.

Attitudes

As was pointed out in Chapter III, the instruments for the measurement of attitudes are limited in scope.

Four questions on the "Psychology Questionnaire" were designed to examine aspects of the students' attitudes toward the programmed material. Since all of the variables except frame size are constant for all versions, differences in responding among treatment groups should reflect the influence of frame size. Of course, previous interests in psychology and other possible individual differences would affect the responses also.

Table 20 reports responses to the question, "Would you like to read more about psychology?" The question assumes that if the material is enjoy-

Table 20

Responses to the Question:
"Would you like to read more about psychology?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	5	11	11	9	9	9	17	71
	%	55.6	91.7	68.8	64.3	56.3	81.8	94.4	71.9
NO	#	4	1	5	5	7	2	1	25
	%	44.4	8.3	31.2	35.7	43.7	18.2	5.6	28.1
Total[a]		9	12	16	14	16	11	18	96

[a] Total cell frequencies differ slightly from those previously stated because questionnaires for two students are not available.

able, the student will want to read more. The expected frequencies (based on the marginal total) in seven of the 14 cells in Table 20 are less than five which prevents the calculation of a χ^2. The percentages of "Yes" responses vary from a low of 55.6 for version 1 to a high of 94.4 for version 7; however, the relationship is not smooth between versions 1 and 7. If version 1 is ignored, the relationship between the percentage of "Yes" responses and versions is curvilinear going from an initial high for version 2 to a low for version 5 and then back up for versions 6 and 7. Unfortunately, little can be concluded from the data in Table 20 since the three highest percentages are for versions 2, 6, and 7. While a case could be made for 6 and 7 differing from the other versions, it is difficult to explain the high response in version 2.

Table 21

Responses to the Question:
"Would you rather read a regular textbook on psychology than have it taught by the materials you have just completed?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	4	7	6	7	8	6	11	49
	%	44.4	63.6	37.5	50.0	50.0	66.7	66.7	53.3
NO	#	5	4	10	7	8	3	6	43
	%	55.6	36.4	62.5	50.0	50.0	33.3	35.3	46.3
Total[a]		9	11	16	14	16	9	17	92

[a] Total of all frequencies differ slightly from those previously stated because questionnaires for two students are not available and because some students did not respond to this question.

Table 21 presents the data for the question, "Would you rather read a regular test book on psychology than have it taught by the materials you have just completed?" Four of the cells (29%) in this table have expected frequencies below 5 which prevents the use of χ^2 to determine if non-chance relationships exist. As with the previous question, there is no clear trend to these responses.

Table 22 contains the responses to the question, "Did you like the way psychology was taught in the materials you have just completed?" Three or about 21% of the cells in this table have expected frequencies below 5. This figure is borderline for χ^2, so a χ^2 was calculated. The obtained

Table 22

Responses to the Question:
"Did you like the way psychology was taught
in the materials you have just completed?"

Response		Version 1	2	3	4	5	6	7	Total
YES	#	5	3	8	8	7	4	7	42
	%	55.6	27.3	50.0	61.5	44.7	36.4	38.8	43.6
NO	#	4	8	8	5	9	7	11	52
	%	44.4	72.7	50.0	38.5	55.3	63.6	61.2	56.4
Total[a]		9	11	16	13	16	11	18	94

$\chi^2 = 4.02$, df = 6, $p < .40$

[a] Total of all frequencies differ slightly from those previously stated because questionnaires for two students are not available and because some students did not respond to this question.

value is 4.02 (df = 6) which is not significant ($p > .60$). No trend is discernable in the percentages of responses across versions; there are more reversals here than in either of the two previous questions.

Responses to the question, "Would you like to become a psychologist?" are reported in Table 23. Half of the cells in this table have expected frequencies of less than 5, which again prevents the use of χ^2. While there is no apparent trend across the first 5 versions, versions 6 and 7 have a higher percentage of "Yes" responses than do the others. The observed frequencies for versions 1 through 5 were combined and those for 6 and 7 were combined and a χ^2 was calculated for the resultant 2 x 2 contingency table. The

Table 23

Responses to the Question:
"Would you like to become a psychologist?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	1	1	3	1	3	4	7	20
	%	11.1	8.3	18.8	7.1	18.8	40.0	38.9	21.1
NO	#	8	11	13	13	13	6	11	75
	%	88.9	91.7	81.2	92.9	81.2	60.0	61.1	78.9
Total[a]		9	12	16	14	16	10	18	95

[a] Total of all frequencies differ slightly from those previously stated because questionnaires for two students are not available and because some students did not respond to this question.

value is highly significant ($\chi^2 = 7.9$, df = 1, $p < .005$). Thus, persons completing versions 6 and 7 indicated that they would like to become psychologists significantly more frequently then those in the other versions. Since no pre-treatment data were collected on this question, it is impossible to ascertain whether this finding is a function of the versions or of some pre-existing interest of the students in these groups. If it is a function of the versions, it indicates that the students reading materials with very few and with no responses are more influenced by the content than those who have more responses. This interpretation should be regarded very cautiously. On the other hand, if it indicates a preexistant interest in psychology, it would suggest that the groups did have initial differences and that the criterion test results may be influenced by differential, preexisting interests. Which of these interpretations best describes the actual situation cannot be determined until the study is replicated.

The questionnaire which was sent to the students after the completion of the study contained several questions which provide information about the student's attitudes. These questions are:

(1) If you were requested to participate in a similar study sometime in the future, would you?

(2) Do you think that a reward should be given to those who participate?

(3) Would you like to see the materials which you read become part of the regular course work at Nova High School?

(4) Did you learn anything about psychology from participating in the study?

Since the students had to complete the questionnaires and mail them back, responses are available for only 63 of the 98 students who completed the study. Additionally 13 of these are anonymous and the version cannot be identified. Thus, only 50 students can be used for comparisons among the versions. This complicates both the analysis and the interpretation of the data because of the reduced N and because of the difficulty of assuming that the questionnaires which were returned are representative of these which were not.

Table 24 presents the data regarding the future participation study. There are too many cells with expected frequency below 5 to calculate a χ^2. No particular trend emerges from the data; most of the students would be willing to participate again.

Table 25 contain the responses regarding the rewards. Again a χ^2

Table 24

Responses to the Question: "If you were requested to participate in a similar study sometime in the future, would you?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	4	2	7	7	5	5	6	36
	%	80.0	50.0	63.6	77.8	83.3	62.5	85.7	72.0
NO	#	0	1	2	0	0	1	0	54
	%	0.0	25.0	18.2	0.0	0.0	12.5	0.0	28.0
DON'T KNOW	#	1	1	2	2	1	2	1	10
	%	20.0	25.0	18.2	22.2	16.7	25.0	14.3	20.0
Total		5	4	11	9	6	8	7	50

cannot be calculated and no clear trend is detectable. Table 26 gives the responses for the question pertaining to the materials becoming part of regular course. A χ^2 cannot be calculated because of the low expected frequencies and no relationship to versions is apparent. The last question is summarized in Table 27. "Yes" responses are very high for every group indicating that almost all the students responding state that they did learn some psychology in the study.

The failure to produce more meaningful results regarding attitudes is primarily due to the limited nature of the instruments. This is unfortunate and should be corrected in future studies.

Table 25

Responses to the Question:
"Do you think that a reward should
be given to those who participate?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	3	0	6	3	3	4	5	24
	%	60.0	0.0	54.5	33.3	50.0	50.0	71.4	48.0
NO	#	2	4	5	6	3	4	2	25
	%	40.0	100.0	45.5	67.3	50.0	50.0	28.6	52.0
Total		5	4	11	9	6	8	7	50

Table 26

Responses to the Question:
"Would you like to see the materials which you read become
part of the regular course work at Nova High School?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	5	0	7	7	4	3	5	31
	%	100.0	0.0	63.6	77.8	66.7	37.5	71.4	62.0
NO	#	0	4	4	2	2	5	2	19
	%	0.0	100.0	36.4	22.2	33.3	62.5	18.6	38.0
Total			5	4	11	9	6	8	50

Table 27

Responses to the Question:
"Did you learn anything about psychology
from participating in the study?"

		Version							
Response		1	2	3	4	5	6	7	Total
YES	#	5	3	9	9	5	7	6	45
	%	100.0	75.0	90.0	100.0	83.3	87.5	85.7	90.0
NO	#	0	1	1	0	1	1	1	5
	%	0.0	25.0	10.0	0.0	16.7	12.5	14.3	10.0
Total		5	4	10	9	6	8	7	50

CHAPTER V: DISCUSSION AND CONCLUSIONS

The most important outcome of the study was the failure to obtain statistical significance among the treatment groups on the criterion test. This suggests that frame size (i.e., the frequency of responding) does not make a difference on the students' mean performance, and that individual differences are the most important determinant of success, holding the other material variables constant. If the thesis that programmed instruction requires overt responding at very frequent intervals is a valid one, then there should have been significant differences among the means on the treatment groups.

It must be remembered that this particular study has not yet been replicated and, therefore, its findings must be regarded as tentative at this point. While the first assumption is that frequency of responding makes no difference, the observed phenomena could have other explanations such as the frequently ignored Type II error where an invalid null hypothesis is not rejected. Further, the criterion test may not be adequately measuring the students' knowledge or it may be unable to differentiate among the versions, even though differences may exist. Although the treatments did not differ on the independent variables studied, there are, of course, many independent variables which were not studied. It is possible that these variables were operative in such a way to negate differences among treatment means. Further, there was not a test for long-term retention, consequently the different treatments may produce differential results over a period of time. Of course, long-term retention is usually the goal of classroom learning and if one treatment did produce more permanent results than another, this is a more important finding than the effects on immediate retention. Consequently, the study should be repeated in the very near future with different samples of students and with different programmed materials. If the findings do withstand replication, it will certainly be a milestone in programmed instruction. In that case, the other variables, such as redundancy and discrete steps, must be analyzed to determine what, if any, the important variables of programmed material are.

An important limitation of the study was the absence of contingencies on the outcome of the criterion test. Rather, responding on the test was a free operant situation in which the students were neither extrinsically punished nor extrinsically rewarded for their performance. The existence of a contingency on the test scores--such as money for a certain obtained score--might force differences in the student's performance among the versions that were not present in the absence of a contingency. Future replications of the study should manipulate criterion test contingencies to determine if they will produce different results.

It should be made clear that the failure of the study to find significant difference among group means on the criterion test does not imply that a particular individual will learn equally well regardless of the frame size or frequency of responding. Programmed learning has frequently been regarded as a universal panacea which will benefit all people equally well. This is probably a naive position, considering all we know about the differences among individuals and the little we know about how humans acquire knowledge. It is difficult to conceive of a unique media or method of instruction working equally well with all people and it is extremely unlikely that such a universal method will ever be found, unless it comes from fields such as psychopharmacology or psychoneurology. Thus, the important aspect of the present study was not to cast doubt upon the necessity for short frames with high response frequency per se, but rather to cast doubt upon this as a universal method. It cannot be determined from the data in the present study whether any particular individual will work better with one method than he does with another method. However, this is certainly a possibility and it is very definitely a topic for further research. While it may very well be that only a few, if any, individuals work at their optimum level with a high frequency of responding, the possibility that many individuals work best with a high frequency of responding is not contradicted by the study.

One of the potential uses of programmed instruction is to aid in individualizing instruction where each person is enabled to learn at his own optimal level within the limits of the available resources. It has been recognized that the conventional classroom containing 30 or 40 students, taught by a single teacher, is not necessarily the best method for imparting knowledge. However, to substitute some other method across the board to all students would be equally bad. Instead, researchers need to provide information which can aid in the establishment of guidelines for the optimal implementation of instructional materials. At present, it is not feasible to study every individual to ascertain the method by which he learns best, although with the aid of computer-based instruction and other forms of computer assistance in the instructional process, such a procedure may eventually be possible. However, until that time, a feasible path for the researcher to follow is to identify various sub-groups or clusters of individuals who have in common the characteristic of optimal learning in the same general way. For example, one group might be identified which learns best using Crowder's auto-tutors for physics, while another group might be identified which learn physics best by the lecture method. It is assumed, of course, that the variables associated with the methods by which each group learns best can be identified and measured. Thus, an important line of research to follow studies such as the present one, is to determine the characteristics of groups, if any, which learn best by programmed instructional materials of each of the versions presented here and by other possible versions.

If a particular group or cluster of students is identified who does succeed best with version 1 of the materials as presented in this study using psychology as a subject matter, then other subject matters should be also investigated with that group to determine if version 1 works equally well regardless of content, or if it interacts significantly with the content. Then, if predictive variables can be identified in that group or cluster, similar groups or clusters can be sampled elsewhere and the studies can be replicated to see if the assumptions hold. In this systematic fashion, researchers can proceed to establish some basic constructs for a theory of teaching or a theory of instruction.

Thus, the present study opens up many avenues of related research; however, the first and most important line of research is the replication of the present study. The findings of no study can be considered definite unless they have been replicated several times without contradictory results. After the replications, and if the findings are maintained, it will be possible to move into other related areas, such as those discussed above.

It would appear that not only does responding or not responding fail to make a difference on the criterion test, but that responding requires significantly longer completion times than nonresponding. The findings on completion time are slightly confusing, because the relationship between frame size and completion time appears to be curvilinear. However, this relationship could be merely a phenomenon unique to this particular sample. If, however, the phenomenon of finding the lowest completion time to be for version 5 (where responses were elicited every sixteenth frame) is a valid finding, it might suggest that occasional responding keeps the student working at a faster pace than either greater or less responding. This might be because the responding is frequent enough to keep his attention, but is not so frequent that it becomes unduly monotonous. However, at this point, this is only speculation and further research must be conducted to determine if the curvilinear relationship holds for other samples and if it does, the reasons why it occurs.

The data of this study do indicate that a high frequency of responding requires more time to complete the materials than do low frequencies of responding. Every version of the material, except version 2, required significantly less time than did version 1. Additionally, 3 of the versions required significantly less time than did version 2. Thus, if the frame size is not related to the criterion test scores, but does affect completion time, an effective argument is made for not having short frames as defined in the present study.

The study of error rate revealed little except that error rate is a weak predictor of the criterion test used. There is not much need to study error rate by itself. The intrinsic interest in error rate manifest in other types of learning studies is not present here, because the acquisition does not occur over repeated trials. The object of learning continually changes throughout the materials so that determination of the change in error rate over time does not necessarily imply the level of learning. A study could be designed to investigate the possible relationships of changes in error rate to other criterion measures, such as achievement tests. The relationship, if it exists, could be an important factor in designing materials to optimize learning. Such a study, however, could be best carried out through the use of computer-assisted instruction where continuous student records can be maintained.

The failure to find conclusive differences in the students' attitudes is not surprising, due to the limited nature of the instruments. As was explained previously, the inadequacies of the questionnaires are largely due to local restrictions on attitude testing. The main intent of the study was to determine the effect of frame size on cognitive acquisition, which was done. However, the affective variables should not be ignored, since they may be more important than the cognitive ones in the long run.

The significant differences which were found are not clearly due to either the sampling and procedure or to preexisting attitudes. In the case of more low-responding-group students wanting to be psychologists than those in the higher-responding-groups, it is unlikely that the random sampling error would cause that concentration, although the attrition conceivably could have caused it. It would seem most likely that the difference is due to treatments, but this cannot be categorically asserted.

On the other hand, the finding that more students who completed the study returned questionnaires than those who did not complete the study is probably due to preexistent factors which also contributed to the attrition itself. That is, the personality variables which led some people to drop out of the study, most likely led them not to return the questionnaire either. However, this cannot be stated with assurance. The explanation may be that those who did not complete the study were embarrassed by the prospect of further contact with the study and ignored the questionnaire to avoid that embarrassment. Also, there could be other explanations, all of which are speculation with little supportive evidence to go on.

Another consideration in measuring attitudes toward new media and new materials in an experimental situation is that novelty effects and Hawthorne effects may distort the measurement. Confrontation with new materials may influence attitudes in either direction. If the person feels less comfortable with the new materials than he does with more familiar materials, he may degrade them. However, if he has had a chance to become accustomed to the new materials, he may rate them higher. On the other hand, the new materials may fascinate him because they are different and he may rate them higher than if he were more accustomed to them. The Hawthorne effect, of course, would inflate the ratings across all experimental groups. Thus, differences in attitude may, in part, be due to these influences of the situations.

In summary, this study has presented evidence that the Skinnerian tenet of small-sized frames, each with its own elicitation of an overt response may not be valid, since no difference could be statistically detected among the seven treatment groups. Since this finding could be due to extraneous factors influencing the data, it must be replicated - as all studies should be. Further, while the study suggests that students in general do not differ across versions, it does not imply that a particular individual will not do better with one method than with another. The completion time data suggest that occasional responding can decrease the time required to complete the materials. This finding also must be replicated.

APPENDIX A: REASONS FOR ATTRITION

A questionnaire was sent to all of the students who had been selected for the study and who did not finish to determine why they did not finish. The questionnaire listed 6 different reasons for non-completion, plus an "other" reason. The reasons listed on the questionnaire are:

(1) I was absent from school too much of the time.

(2) Regular school work kept me too busy.

(3) My parents did not want me to participate.

(4) I felt that the study was a waste of my time.

(5) I just did not want to participate.

(6) I wanted to finish, but the study was discontinued before I was done.

Reasons 1, 2, 4 and 6 were included because of specific feedback from the students in the study. Reason 3 was included because many of the children whose parents called the researcher after they were notified of the study did not participate. Reason 5 was added to detect if some students felt uneasy about research participation.

Table 28 presents the results of the questionnaires. No one checked reasons 3 and 5 so these have been excluded from the table. In some cases, students checked multiple reasons so that the sums of the frequencies exceed the number of students. Since all of the students did not return the questionnaire, the frequencies listed cannot be regarded as descriptive of all the non-participants. In fact, the people who did not return the questionnaire should probably be considered to have different reasons for not participating from those who did return the questionnaire.

Table 28

Reasons Given by Students Who Did Not Complete the Study

	Completion Status			
		Participation		
	No Participation	Did Not Finish the Materials	Finished Materials but Did Not Take Test	Total
Total in Category	15	36	17	68
Number Responding	9	14 (39%)	9 (53%)	32 (47%)
Reason Given				
Absent from School	1	6	2	9
Too Busy With School Work	6	12	2	20
The Study was a Waste of Time	0	3	0	3
Study was Discontinued Too Soon	0	3	1	4
Other[a]	2	2	5	9
Total Responses[b]	9	26	10	45

[a] Other reasons group into 4 categories: Busy in extracurricular activities (3 students); could not find time to take the test (2 students); did not like the material (2 students); and misunderstood (2 students).

[b] Because multiple responses were permitted, total responses exceeds the number of respondents.

In general, the most common reason given is "too busy with other school work." This is reflective of what one student termed the "end of grading period crush" which conflicted with the terminal days of the study -- a conflict which unfortunately arose due to unforeseeable circumstances which delayed the study. One girl who did not complete the materials wrote: "I was very sorry not to be able to complete the test, but as it took place the last two weeks of the grading period, I was unable to miss the time from my various classes." Another girl wrote that there were "too many notebooks and tests due to continue in study." A boy checked "Regular school work" and wrote "I wanted to complete," while a girl checking the same reason wrote "It isn't because I didn't like it -- it interested me very much." Consequently, it is assumed that had the timing been different, the attrition rate would have been less.

APPENDIX B: PROGRAMMED MATERIAL

Examples of each of the seven versions of material are presented on the following pages. The content is identical for all versions and is from the first four pages of Part I of the materials used.

VERSION 1

	A doctor taps your knee (patellar tendon) with a rubber mallet to test your __________.
reflexes (reflex)	If your reflexes are normal, your leg __________ to the tap on the knee with a slight jerk (the so-called knee jerk).
responds (reacts)	In the knee jerk or patellar-tendon reflex, the kick of the leg is the __________to the tap on the knee.
response (reaction)	The stimulating object used by the doctor to elicit a knee jerk is a __________.
hammer (mallet)	The stimulus which elicits a knee jerk is the _____ delivered by the so-called stimulus object or hammer.
tap (blow)	In the knee jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the _________.
stimulus	An event is explained when its cause is identified. The "cause" or explanation of the knee jerk is technically the __________which elicits it.
stimulus (tap on the knee)	Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus __________ a response.

VERSION 1 (Con't)

elicits	In a reflex, the stimulus and the elicited response occur in a given temporal order; first the stimulus then the ______________.
response	A kick of the leg is ____________by a tap on the patellar-tendon.
elicited	The time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the __________ of the knee-jerk reflex.
latency	The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the _______________.
threshold	If you _________ when something brushes your eye, the blink is a response.
blink	The fraction of a second which elapses between "brushing the eye" and "blink" is the ____________ of the reflex.
latency	In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the __________.
stimulus (tap)	The function of a response corresponds to (is a function of) the ____________of the stimulus which elicits it.

VERSION 1 (Con't)

intensity (magnitude, strength)	A reflex consists of both a stimulus and a response, occurring in that order. The term "reflex" is not synonymous with the single term ____________.
response	When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a _______________.
reflex	If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the ___________.
threshold	In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the __________.
latency	If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the _______.
threshold	The greater the concentration of onion juice (stimulus), the ________ the magnitude of the response.
greater (higher-larger)	When speaking technically, instead of saying onion juice "stimulates" tears; we say onion juice ________ tears.
elicits	A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short _________.

VERSION 1 (Con't)

latency	In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful ________ to the hand.
stimulus	In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the __________ or no response will occur.
threshold	A light flashed into the eye elicits constriction of the pupil. This sequence is called pupillary __________.
reflex	In the pupillary reflex, a flash of light is said to _____________ response.
elicit	In the pupillary reflex, a very bright flash of light elicits a response of greater __________ than a weak flash of light.
magnitude (intensity)	A response and its eliciting stimulus comprise a ___________.
reflex	The softest touch on the surface of the eye needed to elicit a blink, marks the ___________ of the stimulus.
threshold	In any reflex, there is a stimulus which __________ the response.
elicits	

VERSION 2

	A doctor taps your knee (patellar tendon) with a rubber mallet to test your reflexes.
	If your reflexes are normal, your leg________ to the tap on the knee with a slight jerk (the so-called knee jerk).
responds (reacts)	In the knee jerk or patellar-tendon reflex, the kick of the leg is the response to the tap on the knee.
	The stimulating object used by the doctor to elicit a knee jerk is a ________.
hammer (mallet)	The stimulus which elicits a knee jerk is the tap delivered by the so-called stimulus object or hammer.
	In the knee jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the _______.
stimulus	An event is explained when its cause is identified. The "cause" or explanation of the knee jerk is technically the stimulus which elicits it.
	Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus elicits a ___________.

VERSION 2 (Con't)

elicits	In a reflex, the stimulus and the elicited response occur in a given temporal order: first the stimulus then the response.
	A kick of the leg is __________ by a tap on the patellar-tendon.
elicted	The time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the latency of the knee-jerk reflex.
	The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the _______.
threshold	If you blink when something brushes your eye, the blink is a response.
	The fraction of a second which elapses between "brushing the eye" and "blink" is the ___________ of the reflex.
latency	In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the stimulus.
	The magnitude of a response corresponds to (is a function of) the __________ of the stimulus which elicits it.

intensity (magnitude, strength)	A reflex consists of both a stimulus and a response, occurring in that order. The term"reflex" is not synonymous with the single term "response."
	When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a ________.
reflex	If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the threshold.
	In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the __________.
latency	If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the threshold.
	The greater the concentration of onion juice (stimulus), the _________ the magnitude of the response.
greater (higher-larger)	When speaking technically, instead of saying onion juice "stimulates" tears; we say onion juice "elicits" tears.
	A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short ___________.

VERSION 2 (Con't)

latency	In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful stimulus to the hand.
	In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the __________ or no response will occur.
threshold	A light flashed into the eye elicits constriction of the pupil. This sequence is called pupillary reflex.
	In the pupillary reflex, a flash of light is said to ____________ response.
elicit	In the pupillary reflex, a very bright flash of light elicits a response of greater magnitude than a weak flash of light.
	A response and its eliciting stimulus comprise a ____________.
reflex	The softest touch on the surface of the eye needed to elicit a blink, marks the threshold of the stimulus.
	In any reflex, there is a stimulus which __________ the response.
elicits	

VERSION 3

A doctor taps your knee (patellar-tendon) with a rubber mallet to test your reflexes.

If your reflexes are normal, your leg responds to the tap on the knee with a slight jerk (the so-called knee jerk).

In the knee jerk or patellar-tendon reflex, the kick of the leg is the response to the tap on the knee.

The stimulating object used by the doctor to elicit a knee jerk is a ____________.

Hammer
(mallet)

The stimulus which elicits a knee jerk is the tap delivered by the so-called stimulus object or hammer.

In the knee jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the stimulus.

An event is explained when its cause is identified. The "cause" or explanation of the knee jerk is technically the stimulus which elicits it.

Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus _________ a response.

VERSION 3 (Con't)

elicits	In a reflex, the stimulus and the elicited response occur in a given temporal order; first the stimulus then the response.
	A kick of the leg is elicited by a tap on the patellar-tendon.
	The time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the latency of the knee-jerk reflex.
	The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the ________.
threshold	If you blink when something brushes your eye, the blink is a response.
	The fraction of a second which elapses between "brushing the eye" and "blink" is the latency of the reflex.
	In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the stimulus.
	The magnitude of a response corresponds to (is a function of) the ________ of the stimulus which elicits it.

VERSION 3 (Con't)

intensity (magnitude, strength)	A reflex consists of both a stimulus and a response occurring in that order. The term "reflex" is not synonymous with the single term "response."
	When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a reflex.
	If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the threshold.
	In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the ________.
latency	If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the threshold.
	The greater the concentration of onion juice (stimulus), the greater the magnitude of the response.
	When speaking technically, instead of saying onion juice "stimulates" tears; we say onion juice "elicits" tears.
	A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short _________.

VERSION 3 (Con't)

latency

In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful stimulus to the hand.

In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the threshold or no response will occur.

A light flashed into the eye elicits constriction of the pupil. This sequence is called pupillary reflex.

In the pupillary reflex, a flash of light is said to ______________ response.

elicit

In the pupillary reflex, a very bright flash of light elicits a response of greater magnitude than a weak flash of light.

A response and its eliciting stimulus comprise a reflex.

The softest touch on the surface of the eye needed to elicit a blink, marks the threshold of the stimulus.

In any reflex, there is a stimulus which __________ any response.

elicits

VERSION 4

A doctor taps your knee (patellar-tendon) with a rubber mallet to test your reflexes.

If your reflexes are normal, your leg responds to the tap on the knee with a slight jerk (the so-called knee-jerk).

In the knee-jerk or patellar-tendon reflex, the kick of the leg is the response to the tap on the knee.

The stimulating object used by the doctor to elicit a knee-jerk is a hammer.

The stimulus which elicits a knee-jerk is the tap delivered by the so-called stimulus object or hammer.

In the knee-jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the stimulus.

An event is explained when its cause is identified. The "cause" or explanation of the knee-jerk is technically the stimulus which elicits it.

Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus ____________ a response.

VERSION 4 (Con't)

elicits

In a reflex, the stimulus and the elicited response occur in a given temporal order; first the stimulus then the response.

A kick of the leg is elicited by a tap on the patellar-tendon.

The time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the latency of the knee-jerk reflex.

The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the threshold.

If you blink when something brushes your eye, the blink is a response.

The fraction of a second which elapses between "brushing the eye" and "blink" is the latency of the reflex.

In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the stimulus.

The magnitude of a response corresponds to (is a function of) the ________ of the stimulus which elicits it.

VERSION 4 (Con't)

intensity
(magnitude,
strength)

A reflex consists of both a stimulus and a response, occurring in that order. The term "reflex" is not synonymous with the single term "response."

When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a reflex.

If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the threshold.

In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the latency.

If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the threshold.

The greater the concentration of onion juice (stimulus), the greater the magnitude of the response.

When speaking technically, instead of saying onion juice "stimulates" tears; we say onion juice "elicits" tears.

A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short ____________.

VERSION 4 (Con't)

latency

In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful stimulus to the hand.

In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the threshold or no response will occur.

A light flashed into the eye elicits a constriction of the pupil. This sequence is called pupillary reflex.

In the pupillary reflex, a flash of light is said to elicit response.

In the pupillary reflex, a very bright flash of light elicits a response of greater magnitude than a weak flash of light.

A response and its eliciting stimulus comprise a reflex.

The softest touch on the surface of the eye needed to elicit a blink, marks the threshold of the stimulus.

In any reflex, there is a stimulus which __________ the response.

elicits

VERSION 5

A doctor taps your knee (patellar-tendon) with a rubber mallet to test your reflexes.

If your reflexes are normal, your leg responds to the tap on the knee with a slight jerk (the so-called knee-jerk).

In the knee-jerk or patellar-tendon reflex, the kick of the leg is the response to the tap on the knee.

The stimulating object used by the doctor to elicit a knee-jerk is a hammer.

The stimulus which elicits a knee-jerk is the tap delivered by the so-called stimulus object or hammer.

In the knee-jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the stimulus.

An event is explained when its cause is identified. The "cause" or explanation of the knee-jerk is technically the stimulus which elicits it.

Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus elicits a response.

VERSION 5 (Con't)

In a reflex, the stimulus and the elicited response occur in a given temporal order; first the stimulus then the response.

A kick of the leg is elicited by a tap on the patellar-tendon.

A time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the latency of the knee-jerk reflex.

The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the threshold.

If you blink when something brushes your eye, the blink is a response.

The fraction of a second which elapses between "brushing the eye" and "blink" is the latency of the reflex.

In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the stimulus.

The magnitude of a response corresponds to (is a function of) the ________ of the stimulus which elicits it.

VERSION 5 (Con't)

intensity (magnitude, strength)	A reflex consists of both a stimulus and a response, occurring in that order. The term "reflex" is not synonymous with the single term "response."

When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a reflex.

If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the threshold.

In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the latency.

If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the threshold.

The greater the concentration of onion juice (stimulus), the greater the magnitude of the response.

When speaking technically, instead of saying onion juice "stimulates" tears, we say onion juice "elicits" tears.

A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short latency.

VERSION 5 (Con't)

In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful stimulus to the hand.

In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the threshold or no response will occur.

A light flashed into the eye elicits constriction of the pupil. This sequence is called pupillary reflex.

In the pupillary reflex, a flash of light is said to elicit response.

In the pupillary reflex, a very bright flash of light elicits a response of greater magnitude than a weak flash of light.

A response and its eliciting stimulus comprise a reflex.

The softest touch on the surface of the eye needed to elicit a blink, marks the threshold of the stimulus.

In any reflex, there is a stimulus which _________ the response.

elicits

VERSION 6

A doctor taps your knee (patellar-tendon) with a rubber mallet to test your reflexes.

If your reflexes are normal, your leg responds to the tap on the knee with a slight jerk (the so-called knee-jerk).

In the knee jerk or patellar-tendon reflex, the kick of the leg is the response to the tap on the knee.

The stimulating object used by the doctor to elicit a knee-jerk is a hammer.

The stimulus which elicits a knee-jerk is the tap delivered by the so-called stimulus object or hammer.

In the knee-jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the stimulus.

An event is explained when its cause is identified. The "cause" or explanation of the knee-jerk is technically the stimulus which elicits it.

Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus elicits a response.

VERSION 6 (Con't)

In a reflex, the stimulus and the elicited response occur in a given temporal order; first the stimulus then the response.

A kick of the leg is elicited by a tap on the patellar-tendon.

The time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the latency of the knee-jerk reflex.

The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the threshold.

If you blink when something brushes your eye, the blink is a response.

The fraction of a second which elapses between "brushing the eye" and "blink" is the latency of the reflex.

In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the stimulus.

The magnitude of a response corresponds to (is a function of) the intensity of the stimulus which elicits it.

VERSION 6 (Con't)

A reflex consists of both a stimulus and a response occurring in that order. The term "reflex" is not synonymous with the single term "response."

When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a reflex.

If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the threshold.

In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the latency.

If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the threshold.

The greater the concentration of onion juice (stimulus), the greater the magnitude of the response.

When speaking technically, instead of saying onion juice "stimulates" tears; we say onion juice "elicits" tears.

A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short latency.

VERSION 6 (Con't)

In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful stimulus to the hand.

In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the threshold or no response will occur.

A light flashed into the eye elicits a constriction of the pupil. This sequence is called pupillary reflex.

In the pupillary reflex, a flash of light is said to elicit a response.

In the pupillary reflex, a very bright flash of light elicits a response of greater magnitude than a weak flash of light.

A response and its eliciting stimulus comprise a reflex.

The softest touch on the surface of the eye needed to elicit a blink, marks the threshold of the stimulus.

In any reflex, there is a stimulus which __________ the response.

VERSION 7

A doctor taps your knee (patellar-tendon) with a rubber mallet to test your reflexes.

If your reflexes are normal, your leg responds to the tap on the knee with a slight jerk (the so-called knee jerk).

In the knee jerk or patellar-tendon reflex, the kick of the leg is the response to the tap on the knee.

The stimulating object used by the doctor to elicit a knee jerk is a hammer.

The stimulus which elicits a knee jerk is the tap delivered by the so-called stimulus object or hammer.

In the knee jerk reflex, we call the rubber hammer the stimulus object and the tap or blow the stimulus.

An event is explained when its cause is identified. The "cause" or explanation of the knee jerk is technically the stimulus which elicits it.

Technically speaking, a reflex involves an eliciting stimulus in a process called an elicitation. A stimulus elicits a response.

VERSION 7 (Con't)

In a reflex, the stimulus and the elicited response occur in a given temporal order; first the stimulus then the response.

A kick of the leg is elicited by a tap on the patellar-tendon.

The time which elapses between the onset of the stimulus and the onset of the response is called the latency. Thus the time between the tap and the kick is the latency of the knee-jerk reflex.

The weakest stimulus sufficient to elicit a response marks the threshold of the reflex. A tap on the knee will not elicit a kick if it is below the threshold.

If you blink when something brushes your eye, the blink is a response.

The fraction of a second which elapses between "brushing the eye" and "blink" is the latency of the reflex.

In the patellar-tendon reflex, a forceful tap elicits a strong kick; a tap barely above the threshold elicits a weak kick. Magnitude of response depends on the intensity of the stimulus.

The magnitude of a response corresponds to (is a function of) the intensity of the stimulus which elicits it.

VERSION 7 (Con't)

A reflex consists of both a stimulus and a response occurring in that order. The term "reflex" is not synonymous with the single term "response."

When a person is startled by a loud noise, his sudden movement is his response to the noise which has acted as a stimulus. The two together are called a reflex.

If a sip of a very weak lemonade does not cause salivation, the stimulus is said to be below the threshold.

In the salivary reflex, the stimulus (food) precedes the responses (secretion of saliva) by an interval of time called the latency.

If an actress uses onion juice on her handkerchief to elicit tears during an emotional scene, she must use enough to exceed the threshold.

The greater the concentration of onion juice (stimulus), the greater the magnitude of the response.

When speaking technically, instead of saying onion juice "stimulates" tears; we say onion juice "elicits" tears.

A child touching a hot surface withdraws his hand quickly. The word "quickly" suggests that the response has a short latency.

VERSION 7 (Con't)

In withdrawing the hand from the hot surface, arm movement is a response which is elicited by a painful stimulus to the hand.

In the hand-withdrawal reflex, the stimulus must be intense enough to exceed the threshold or no response will occur.

A light flashed into the eye elicits a constriction of the pupil. This sequence is called pupillary reflex.

In the pupillary reflex, a flash of light is said to elicit response.

In the pupillary reflex, a very bright flash of light elicits a response of greater magnitude than a weak flash of light.

A response and its eliciting stimulus comprise a reflex.

The softest touch on the surface of the eye needed to elicit a blink, marks the threshold of the stimulus.

In any reflex, there is a stimulus which elicits the response.

APPENDIX C: INSTRUMENTS

PSYCHOLOGICAL ACHIEVEMENT TEST

FORM A

- Do not write on this test booklet -

- Write only on the answer cards provided -

- Use only the special pencil -

INSTRUCTIONS:

1. This is a multiple choice test in psychology. There is no time limit for this test. Most students will finish in 45 to 60 minutes.

2. Complete both sides of the yellow STUDENT DATA CARD following the instructions on the card.

3. Write your name and grade as indicated on the front of the white TEST ANSWER CARD.

4. Fill in the bubble labeled "Part I" and answer questions 1-50 on the front of the TEST ANSWER CARD. Fill in the bubble labeled "Part II" and answer questions 51-64 on the back of the TEST ANSWER CARD.

5. If you have any questions about how to take this examination, ask the monitor.

- DO NOT WRITE ON THIS TEST BOOKLET -

Prepared by J. M. Flynn, Nova University, 1968

1. During an air attack the odor of a bomb shelter may be paired with the stimuli which elicits fear during the attack. If so, this will be an example of

    1) arranged operant conditioning.
    2) arranged respondent conditioning.
    3) accidental operant conditioning.
    4) accidental respondent conditioning.

2. Head scratching, pencil chewing and table tapping by a student while he is solving a problem are most likely to be a form of

    1) unconditioned problem solving behavior,
    2) superstitious problem solving behavior.
    3) generalized problem solving behavior.
    4) discriminative problem solving behavior.

3. A thirsty man who sees a drinking fountain will probably go to the fountain because

    1) walking to the fountain has been reinforced in the past when he was thirsty.
    2) walking to the fountain has been extinguished in the past when he was thirsty.
    3) walking to the fountain is an unconditioned response.
    4) the fountain is an unconditioned stimulus.

4. A green key is an $S\Delta$ for pecking by a pigeon, and a red key is an S^D in a fixed ratio schedule of 25 responses. After the key turns green,

    1) no responses are reinforced.
    2) all responses are reinforced.
    3) the twenty-fifth response is reinforced.
    4) The twenty-sixth response is reinforced.

5. In a typical operant conditioning experiment, the experimenter arranges a mechanical or electrical device to

    1) reinforce any response.
    2) reinforce a specific response.
    3) Reinforce occasional responses regardless of their nature.
    4) reinforce any frequently occurring response.

6. In the usual operant experiment, bending down to the magazine makes the food visible and the pigeon seizes the food with its beak. Which of the following statements most correctly describes the situation?

    1) With respect to seizing food, the sight of food is a $S\Delta$.
    2) With respect to bending down, the sight of food is a reinforcer.
    3) With respect to seizing food, the sigh of food is a S^D.
    4) Statements 2. & 3. above are correct, but not statement 1.

7. The highest rate of responding is generated by which of the following schedules of reinforcement

   1) variable-ratio
   2) fixed-interval
   3) variable-interval
   4) fixed-ratio

8. A chimpanzee drops coins into a vending machine which delivers peanuts. The peanuts are

   1) conditioned reinforcers
   2) secondary reinforcers
   3) unconditioned reinforcers
   4) accidental reinforcers

9. If a drill instructor says "right face!" and all of his men turn to the right, it can be said their behavior

   1) is under the control of a verbal stimulus.
   2) is an unconditioned response.
   3) is reinforced on a fixed-ratio schedule.
   4) demonstrates complete stimulus generalization.

10. A particular schedule of reinforcement always has 325 responses between reinforcements. This schedule is called

    1) variable-interval schedule
    2) variable-ratio schedule
    3) fixed-interval schedule
    4) fixed-ratio schedule

11. In teaching a child to copy a picture, if we reinforce only a perfect drawing, we can most likely expect which one of the following results?

    1) His drawing behavior will be extinguished.
    2) His drawing behavior will rapidly improve.
    3) His drawing behavior will be slowly improved.
    4) While his drawing behavior will not be extinguished, but it will not improve.

12. Smooth muscle fibers are found in which one of the following:

    1) the arms
    2) the stomach
    3) the neck
    4) the ankle

13. In the standard experimental box the click of the food magazine is an effective reinforcer for pecking because

    1) the click occurs instantly but there is a short delay before reinforcement with food.
    2) food occurs instantly but there is a short delay before reinforcement with the click.
    3) reinforcement by the click and by the food both occur instantly.
    4) reinforcement with the click and the food are both delayed.

14. The technique of reinforcing a response in the presence of one stimulus and not reinforcing it in the presence of another stimulus is used to establish

    1) an adaption
    2) a superstition
    3) a discrimination
    4) a generalized reinforcer

15. The term which best describes the situation in which some responses are reinforced while some rather similar responses are not reinforced is

    1) continuous reinforcement
    2) differential reinforcement
    3) accidental reinforcement
    4) generalized reinforcement

16. A child has a temper tantrum and screams for candy. The mother gives the candy to the child and the tantrum ceases. Which one of the following statements best describes this situation?

    1) Termination of the tantrum negatively reinforces the mother's response of handing candy to the child.
    2) The termination of the tantrum positively reinforces the mother's response of handing candy to the child.
    3) Handing candy to the child negatively reinforces the tantrum.
    4) Neither the mother's response of handing candy to the child nor the child's tantrum are reinforced at all.

17. If only red apples are always sweet and edible and green apples are never sweet and edible, which of the following statements is correct for the response of picking and eating an apple?

    1) The "redness" or "greenness" of an apple is an S^D.
    2) The "redness" or "greenness" of an apple is an S^Δ.
    3) The "redness" of an apple is an S^Δ and the "greenness" of an apple is an S^D.
    4) The "redness" of an apple is an S^D and the "greenness" of an apple is an S^Δ.

18. Vomiting due to chemical irritants in the stomach is:

    1) a conditioned response.
    2) an unconditioned response.
    3) a conditioned stimulus.
    4) an unconditioned stimulus.

19. In a chain of behavior, the presence of food

    1) can be a discriminative stimulus, but not a reinforcing stimulus.
    2) can be a reinforcing stimulus, but not a discriminative stimulus.
    3) can be both a discriminative and reinforcing stimulus.
    4) can be neither a discriminative nor reinforcing stimulus.

20. Which of the following statements is true of continuous reinforcement:

    1) Every response is reinforced and the time required for extinction is long.
    2) Many, but not all, of the responses are reinforced and the time required for extinction is short.
    3) Every response is reinforced and the time required for extinction is short.
    4) Many, but not all, of the responses are reinforced and the time required for extinction is long.

21. In a typical operant conditioning experiment, a pigeon might be placed in a box which has a key for pecking and a food magazine. To condition operant behavior the apparatus

    1) pairs a conditioned stimulus with the presentation of food in the magazine.
    2) reinforces any response which occurs by presenting the animal with food.
    3) elicits specific responses by presentation of food.
    4) reinforces specific responses when they are emitted.

22. Even though reinforcement can only occur after a response, the shortest interval in a variable-interval schedule

    1) may not be any smaller than the average interval.
    2) may be so small that one reinforcement occurs immediately after another reinforcement.
    3) must be at least one minute in length.
    4) is the same length as the longest interval because all of the intervals are the same length.

23. If a pigeon's key pecking response is not followed by reinforcement,

key pecking responses will occur

1) less frequently in the future than if the response had been followed by reinforcement.
2) more frequently in the future than if the response had been followed by reinforcement.
3) as frequently in the future as if the response had been followed by reinforcement.
4) no statement can be made about future responses since the frequency depends upon the particular animal.

Read the following before answering questions 24 and 25:

In a standard respondent conditioning experiment a tone was paired with a meat powder and the saliva flow of the dog was measured. There were 60 trials. On trials 1, 12, 24, 36, 48 and 60, the tone was presented alone (i.e., meat powder was not used). The table below shows the amount of saliva flow and the time between onset of tone and salivation on each of these test trials.

Trial No.	Drops of Saliva	Time Between Onset of Tone and Salivation (in seconds)
1	0	-
12	8	15
24	27	7
36	68	3
48	72	2
60	71	2

24. On Trial 12, the tone alone was presented and 8 drops of saliva flowed. This is evidence that the tone was already:

    1) a conditioned response
    2) a conditioned stimulus
    3) an unconditioned response
    4) an unconditioned stimulus

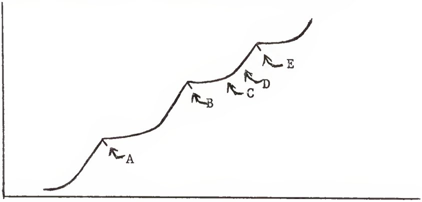
25. From this experiment, we can conclude that as the number of trials in which the conditioned and unconditioned stimuli are paired increases:

    1) the latency of the conditioned reflex increases and the magnitude of the conditioned response increases until both reach a limit.
    2) the latency of the conditioned reflex decreases and the magnitude of the conditioned response decreases until both reach a limit.
    3) the latency of the conditioned reflex decreases and the magnitude of the conditioned response increases until both reach a limit.

4) The latency of the conditioned reflex increases and the magnitude of the conditioned response decreases until both reach a limit.

Read the following figure and use it for questions 26 and 27.

RECORD OF A PIGEON ON A FIXED INTERVAL SCHEDULE



26. At which of the following points in the above figure were the responses reinforced?

1) A, B, and C
2) A, B, and D
3) B, D, and E
4) A, B, and E

27. At which of the following intervals in the above figure is the rate the lowest?

1) between C and D
2) between D and E
3) between B and C
4) between A and B

28. If a cigarette lighter fails to light and the person stops trying to light it after a very few attempts, we can most probably assume that his previous lighting responses were reinforced on a

1) continuous schedule
2) variable-interval schedule
3) variable-ratio schedule
4) fixed-ratio schedule

29. If in teaching a student to throw a javelin the track coach is satisfied with every throw no matter how bad it may be, he is using which of the following techniques.

1) he is using successive approximation, but he is *not* using differential reinforcement.
2) he is *not* using successive approximation, but he is using differential reinforcement.
3) he is using successive approximation and he is using differential reinforcement.
4) he is *not* using successive approximation and he is *not* using differential reinforcement.

30. Which one of the following statements about reinforcement is correct?

1) Positive reinforcement increases the rate of response while negative reinforcement decreases the rate of response.
2) Both positive and negative reinforcement increase the rate of response.
3) Positive reinforcement decreases the rate of response and negative reinforcement increases the rate of responses.
4) Both positive and negative reinforcement decrease the rate of response.

31. When reinforcements are *never* close together in time, which of the following situations would most likely occur immediately after reinforcement?

1) Rapid responding
2) A few responses followed by a long pause.
3) A single response followed by a long pause.
4) A pause in responding.

32. In shaping animal behavior such as that of a dog the first step is often to establish a sound such as a click as a reinforcer by repeatedly pairing it with food. One reason for using an auditory stimulus instead of food in training the dog is

1) the auditory stimulus is a primary reinforcer whereas the food is not.
2) the auditory stimulus affects the dog immediately whereas the food may not.
3) the auditory stimulus is a generalized reinforcer whereas the food is not.
4) the click can be used for continuous reinforcement whereas the food can only be used on a ratio schedule.

33. If we consistently get no answer when we dial a number on the telephone we stop dialing. This process is called

1) negative reinforcement
2) positive reinforcement
3) extinction
4) adaption

34. When you telephone a friend not at home, your reinforcement by getting an answer is on

1) a ratio-schedule.
2) an interval-schedule.
3) a continuous reinforcement schedule.
4) no particular reinforcement schedule.

35. Which of the following methods will most likely be used to shape a complex skill?

1) generalized reinforcement of unconditioned responses
2) pairing of conditioned stimuli with unconditioned stimuli
3) differential reinforcement of successive approximations to the final behavior
4) continuous reinforcement of generalized behavior through the use of secondary reinforcers

36. The adage "practice makes perfect" is probably true only if some product of one's behavior provides

1) generalized reinforcement.
2) primary reinforcement.
3) differential reinforcement.
4) accidental reinforcement.

37. A dog has been conditioned to salivate at the sound of a tuning fork. Later, the experimenter wants to extinguish the conditioned response. Which of the following methods would lead to extinction:

1) Repeated presentation of the tuning fork with food powder.
2) Repeated presentation of food powder without the tuning fork.
3) Repeated presentation of the tuning fork without meat powder.
4) Repeated presentation of meat powder occasionally paired with the tuning fork.

38. Which of the following is true for a multiple schedule?

1) Behavior in a multiple schedule can only be reinforced by primary reinforcers.
2) A multiple schedule can consist only of a fixed interval schedule alternating with a fixed ratio schedule.
3) In a multiple schedule, behavior comes under stimulus control.
4) Reinforcement in a multiple schedule can only be by conditioned reinforcers.

39. A response is said to have become more probable if under controlled conditions

1) it is observed to occur more frequently than it did at some previous time.
2) it is observed to occur less frequently than at some previous time.
3) its latency is observed to increase.
4) its magnitude is observed to increase.

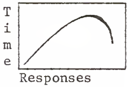
40. One way in which to train a dog not to sit still is by withholding reinforcement until the dog moves. The term which best describes this technique is called

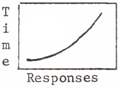
1) accidental reinforcement
2) primary reinforcement
3) continuous reinforcement
4) differential reinforcement

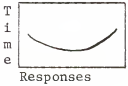
41. If a pigeon is placed in an experimental box for the first time, which of the following pecking response patterns is most likely to occur before the pigeon becomes accustomed to the situation?

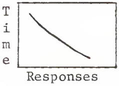
1) Relatively long pauses between responses.
2) extremely short pauses between responses.
3) alternating long and short pauses between responses.
4) none of the above.

42. If the time between pecking responses for a pigeon placed in an experimental box for the first time and under a schedule of continuous reinforcement were plotted, which of the following figures would most nearly resemble the resultant graph?

1)


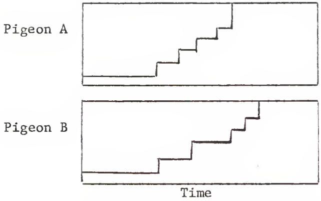
3)


2)


4)


Refer to the following figures for questions 43 and 44.

The figures shown are the cumulative records of conditioned key pecking by two pigeons. All responses were reinforced.



43. In Figure A, Pigeon A reached his maximal rate of responding after

    1) 1 reinforcement
    2) 4 reinforcements
    3) 8 reinforcements
    4) 11 reinforcements

44. The record for Pigeon B differs from that for Pigeon A mainly in the fact that

    1) the maximal rate of responding was significantly faster.
    2) the maximal rate of responding was significantly slower.
    3) a longer time elapsed between the first, second and third responses.
    4) the responses for Pigeon B occur more frequently.

45. Which of the following are a direct result of actions of the glands:

    1) vomiting
    2) sweating
    3) blushing
    4) breathing

46. A bowler may snap his fingers just as the ball hits the pins. If he gets a "strike"

    1) his finger snapping will not be reinforced because this behavior has no effect on the ball.
    2) his finger snapping will be reinforced even though this behavior has no effect on the ball.
    3) his finger snapping will not be reinforced although this behavior does have an effect on the ball.
    4) his finger snapping will be reinforced because this behavior does have an effect on the ball.

47. "New" forms of behavior can be generated by differentially reinforcing forms which gradually approximate the final form. We call this procedure

    1) shaping the behavior
    2) extinguishing the behavior
    3) adapting the behavior
    4) emitting the behavior

48. When a food deprived pigeon is placed for the very first time in a standard experimental box "emotional" behavior is elicited by the box. This behavior is

    1) conditioned and can be eliminated through extinction.
    2) conditioned and can be eliminated through adaption.
    3) unconditioned and can be eliminated through extinction.
    4) unconditioned and can be eliminated through adaption.

49. Sometimes the click made by the empty food magazine in an operant conditioning experiment acquires the power to reinforce the key pecking response of the pigeon. In this case we will call the sound of the magazine

    1) an unconditioned reinforcer
    2) a primary reinforcer
    3) a conditioned reinforcer
    4) a generalized reinforcer

50. Stimuli which were never present at reinforcement

    1) will never acquire control over the response.
    2) may acquire control over the response through accidental reinforcement.
    3) frequently acquire control over the response although practically nothing is known about the conditions under which they acquire this control.
    4) may acquire control over the response if a similar stimulus has been present during reinforcement.

51. One way in which to extinguish behavior you do not want another person to emit is by

    1) showing them affection when they emit the behavior
    2) withholding approval when they do not emit the behavior
    3) using a generalized reinforcer when they emit the behavior
    4) withholding approval when they emit the behavior

52. A worker in an electronics factory is paid $3 for every lot of 15 tuning knobs which his machine produces. Since he is reinforced only when he finishes a lot the reinforcement is on

    1) a variable-ratio schedule
    2) a fixed-interval schedule
    3) a continuous-reinforcement schedule
    4) a fixed-ratio schedule

53. When reinforcement is discontinued after a response has been maintained on a variable interval schedule, responses

    1) are emitted at a gradually declining rate for a long period of time.
    2) are emitted at a rapidly declining rate for only a short period of time.
    3) are emitted at a constantly high rate for an indefinite period of time.
    4) will cease to be emitted almost immediately.

54. Which of the following techniques will effectively prevent unwanted conditioned behavior?

    1) Condition some incompatible behavior.
    2) Withhold reinforcement occasionally when the behavior is emitted.
    3) Withhold reinforcement every time the behavior is emitted.
    4) Both 1 and 3 will effectively prevent the behavior but not 2.

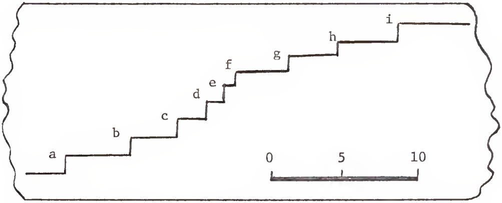
55. Frequently words like "bad" and "wrong" elicit emotional responses in a person. These responses are:

    1) unconditioned responses.
    2) conditioned responses.
    3) sometimes unconditioned responses and other times conditioned responses.
    4) neither unconditioned nor conditioned responses.

56. A reinforcement schedule which provides reinforcements after a set number of responses is called

    1) a variable-ratio schedule
    2) a fixed-ratio schedule
    3) a variable-interval schedule
    4) a fixed-interval schedule

Refer to the following figure for questions 57 and 58. The figure is the cumulative record of keypecking responses by a pigeon.



57. Between which two positions in the above figure did the animal pause without responding for about five minutes?

    1) a and b
    2) b and c
    3) c and d
    4) e and f

58. In the above figure three responses were made fairly close together. Which of the following letters identifies these three responses?

    1) a, b, and c
    2) c, d, and e
    3) b, c, and d
    4) d, e, and f

59. If a conditioned reinforcer is paired with several different uncinditioned reinforcers, it becomes

    1) an accidental reinforcer
    2) an unconditioned reinforcer
    3) a generalized reinforcer
    4) a primary reinforcer

60. Which of the following statements is the best psychological interpretation of the question "what does he see in her?"

    1) "How does he reinforce her courting behavior?"
    2) "How does she reinforce his courting behavior?"
    3) "How does she extinguish his courting behavior?"
    4) "How does he extinguish her courting behavior?"

61. When a student is studying he may persistently scratch his head and chew on his pencil. Frequently these behaviors are conditioned superstitious operants resulting from

    1) differential reinforcements of successive approximations to a final behavior.
    2) the pairing of unconditioned stimuli with conditioned stimuli.
    3) accidental contingencies of reinforcement.
    4) continuous reinforcement of high probability behavior.

62. If the child who first calls all books"green" is not reinforced for saying "green" to books of other colors, he will soon call only green books "green". Thus, books of other colors

    1) come to provide S^D's for the response of saying "green".
    2) become $S\Delta$ for the response of saying other colors.
    3) come to provide $S\Delta$'s for the response of saying "green".
    4) become unconditioned S^D for eliciting the response of saying "green".

63. Which of the various schedules of reinforcement do the following statements describe: There is a pause after reinforcement. Responding then resumes and is positively accelerated until it reaches a terminal steady rate before next reinforcement.

    1) fixed-interval
    2) variable-ratio
    3) variable-interval
    4) fixed-ratio

64. A pigeon has been given food each time it pecks a disk. Later when food no longer follows pecking, the rate in emitting the response is observed to decline gradually. This process is called

    1) operant conditioning
    2) extinction
    3) respondent conditioning
    4) adaption

PSYCHOLOGY QUESTIONNAIRE

Please answer each of the following questions by circling either "yes" or "no" as they apply. Answer every question as truthfully as possible.

Yes No 1. Have you ever been taught in school anything about psychology (besides the materials which you have just completed)?

Yes No 2. Have you ever read a book about psychology (besides the materials which you have just completed)?

Yes No 3. Have you ever read a magazine article about psychology?

Yes No 4. Is anyone you know a psychologist?

Yes No 5. Have you ever visited a psychological laboratory?

Yes No 6. Have you ever talked to a psychologist about psychology?

Yes No 7. Have you ever watched a psychologist work?

Yes No 8. Is your father or mother a psychologist?

Yes No 9. Do you ever talk about psychology at home with members of your family?

Yes No 10. Did you know what a psychologist did before you read the materials you just completed?

Yes No 11. Would you like to read more about psychology?

Yes No 12. Would you rather read a regular textbook on psychology than have it taught by the materials you have just completed?

Yes No 13. Did you like the way psychology was taught in the materials you have just completed?

Yes No 14. Would you like to become a psychologist?

NAME (optional) ______________________ Grade ______ Sex _____

Please check the following items which apply:

A. Please check one:

_____ (1) I did not participate at all in the study.
_____ (2) I participated for only a few days in the study.
_____ (3) I participated for most of the time, but I did not finish reading the material.
_____ (4) I finished reading the material, but I did not complete the questionnaire or the test.
_____ (5) I finished the material and the questionnaire, but I did not take the test.
_____ (6) I completed all parts of the study.

Comments (if any):

B. If you did not finish the study please check the reason (or reasons):

_____ (1) I was absent from school too much of the time.
_____ (2) Regular school work kept me too busy.
_____ (3) My parents did not want me to participate.
_____ (4) I felt that the study was a waste of my time.
_____ (5) I just did not want to participate.
_____ (6) I wanted to finish, but the study was discontinued before I was done.
_____ (7) Other (please explain) ______________________________
____

Comments (if any):

C. If you were requested to participate in a similar study sometime in the future, would you?

_____ (1) Yes, I would participate.
_____ (2) No, I would not participate.
_____ (3) I do not know if I would participate or not.

Comments (if any):

D. Some people have suggested that some form of a reward be given to students who do participate in studies such as the one just completed. Such a reward might be an additional priviledge at the school or some kind of a present such as an LP album or a pass to a movie theater.

Do you think that a reward should be given to those who participate?

Yes _____ No _____

If you do, what should it be: ______________________________
____

E. Would you like to see the materials which you read become part of the regular course work at Nova High School?

Yes _______ No _______

F. Did you learn anything about psychology from participating in the study?

Yes _______ No _______

APPENDIX D: DATA

Tables 28 through 34 contain the data used in the study. Table 28 contains the data for students in Version 1 of the material, Table 29 contains the data for Version 2, and so on. The data included are listed below, keyed according to the columnar headings in the tables:

Id.# - An arbitrary number assigned to each student. It is not the school-assigned student number.

Version - The number of the material version as described in Chapter II.

Group - The letter representing the physical grouping of each student used for scheduling.

Sex - 0 represents male and 1 female.

Grade - 0 represents tenth grade and 1 represents eleventh grade.

Time - The time in minutes in which each student completed the materials.

Error - The errors made by each student.

R - the actual number of errors

% - the percentage of errors

T - T-scores based on the mean and standard deviations of the students in each version

SCAT (Perc.) - SCAT Verbal, Quantitative, and Total scores reported in percentiles based on Florida norms.

SCAT (T) - SCAT Verbal, Quantitative, and Total scores in which the percentiles have been converted to normalized T-scores.

Status - The completion status of each of the students according to the following code:

0 - did not start

1 - participated 1 or 2 sessions

2 - participated most of the time, but did not finish

3 - finished reading, but did not take the test

4 - finished all parts of the study

Additional information on the data is presented in the text -- especially in Chapter II.

Table 29

Data for Students in Version 1 of the Material

						Error				SCAT (Perc.)			SCAT (T)			
Id.#	Version	Group	Sex	Grade	Time	R	%	T	Test	Verb.	Quant.	Tot.	Verb.	Quant.	Tot.	Status
101	1	A	0	1	-	-	-	-	-	55	32	46	51	45	49	3
102	1	A	1	1	355	22	3	44	49	98	88	97	71	62	69	4
103	1	A	1	0	-	-	-	-	-	24	10	17	43	37	40	3
104	1	A	0	0	-	-	-	-	-	89	92	94	62	64	66	2
105	1	B	1	1	335	42	5	52	15	08	36	14	36	46	39	4
106	1	B	0	0	260	35	4	49	24	52	74	64	51	56	54	4
107	1	B	1	0	-	-	-	-	-	06	01	02	34	27	29	2
108	1	C	1	1	-	-	-	-	-	69	36	57	55	46	52	2
109	1	C	0	0	335	42	5	52	36	82	28	62	59	44	53	4
110	1	D	0	0	290	21	2	43	49	-	-	-	-	-	-	4
111	1	D	1	0	315	28	3	46	25	75	60	71	57	52	56	4
112	1	D	1	1	425	97	11	77	43	-	-	-	-	-	-	4
113	1	E	0	0	-	-	-	-	-	96	83	94	68	55	66	3
114	1	E	1	0	420	10	1	38	50	77	80	82	57	52	59	4
115	1	E	0	1	400	26	3	45	55	49	77	64	50	50	54	4
116	1	F	1	0	-	-	-	-	-	87	77	87	61	53	61	3
117	1	F	1	1	-	-	-	-	-	75	80	81	57	52	59	2
118	1	F	0	0	335	41	5	52	30	96	88	95	68	55	66	4
119	1	B	0	1	-	-	-	-	-	34	74	51	46	49	50	0
120	1	C	0	1	-	-	-	-	-	16	20	15	40	47	40	0
121	1	E	1	1	-	-	-	-	-	69	44	61	55	51	53	0
122	1	F	0	1	-	-	-	-	-	80	80	84	58	52	60	0
123	1	D	0	1	-	-	-	-	-	27	52	35	44	48	46	0

Table 30

Data for Students in Version 2 of the Material

						Error				SCAT (Perc.)			SCAT (T)			
Id.#	Version	Group	Sex	Grade	Time	R	%	T	Test	Verb.	Quant.	Tot.	Verb.	Quant.	Tot.	Status
124	2	A	0	1	265	6	1	41	42	85	67	81	60	54	59	4
125	2	A	1	1	-	-	-	-	-	47	74	61	49	56	53	3
126	2	A	1	0	-	-	-	-	-	-	-	-	-	-	-	3
127	2	A	0	1	430	42	10	52	50	89	92	94	62	64	66	4
128	2	A	0	0	-	-	-	-	-	82	71	81	59	56	59	2
129	2	B	0	1	-	-	-	-	-	16	17	14	40	40	39	3
130	2	B	1	1	-	-	-	-	-	61	67	66	53	54	54	2
131	2	B	1	0	290	20	5	45	42	72	52	66	56	51	54	4
132	2	B	0	0	-	-	-	-	-	55	10	35	51	37	46	2
133	2	C	1	0	335	50	12	54	37	77	60	73	57	53	56	4
134	2	C	0	0	380	82	19	63	28	85	99	97	60	73	69	4
135	2	C	1	1	-	-	-	-	-	-	-	-	-	-	-	3
136	2	C	0	1	290	41	9	51	28	-	-	-	-	-	-	4
137	2	D	0	0	-	-	-	-	-	66	52	62	54	51	53	2
138	2	D	1	1	225	18	4	44	51	99	88	98	73	62	71	4
139	2	E	1	0	315	25	29	76	20	47	40	44	49	47	48	4
140	2	E	0	0	275	16	4	44	31	96	97	98	68	69	71	4
141	2	E	1	1	370	6	1	41	29	63	74	71	53	56	56	4
142	2	F	1	1	285	12	3	43	48	95	77	91	66	57	63	4
143	2	F	0	1	-	-	-	-	-	58	80	71	52	58	56	2
144	2	F	1	0	215	25	6	46	35	-	-	-	-	-	-	4
145	2	D	0	1	-	-	-	-	-	22	13	35	42	39	46	0
146	2	D	1	0	-	-	-	-	-	58	36	49	52	46	50	0
147	2	E	0	1	-	-	-	-	-	18	32	21	41	45	42	0

Table 31

Data for Students in Version 3 of the Material

Id.#	Version	Group	Sex	Grade	Time	Error R	Error %	Error T	Test	SCAT (Perc.) Verb.	SCAT (Perc.) Quant.	SCAT (Perc.) Tot.	SCAT (Tot.) Verb.	SCAT (Tot.) Quant.	SCAT (Tot.) Tot.	Status
148	3	A	1	1	370	8	4	43	31	44	63	53	48	53	51	4
149	3	A	0	1	-	-	-	-	-	61	56	61	53	52	53	2
150	3	A	1	0	310	49	23	65	18	-	-	-	-	-	-	4
151	3	A	0	0	255	10	5	44	26	36	40	37	46	47	47	4
152	3	B	1	0	310	51	24	66	44	91	90	94	63	63	66	4
153	3	B	0	1	415	18	8	48	24	58	24	44	52	43	48	4
154	3	B	0	0	270	9	4	43	38	77	71	78	57	56	58	4
155	3	C	0	1	-	-	-	-	-	45	38	83	49	47	60	2
156	3	C	1	0	280	39	18	59	44	-	-	-	-	-	-	4
157	3	C	1	1	190	2	1	40	41	85	80	87	60	58	61	4
158	3	C	0	0	320	24	11	51	28	52	77	66	51	57	54	4
159	3	D	1	0	-	-	-	-	-	82	48	71	59	49	56	2
160	3	D	0	0	285	8	4	43	18	89	86	91	62	61	63	4
161	3	D	1	1	-	-	-	-	-	99	77	95	73	57	66	2
162	3	D	0	1	240	5	2	41	34	-	-	-	-	-	-	4
163	3	E	0	0	-	-	-	-	-	18	44	25	41	48	43	2
164	3	E	1	1	255	16	7	47	40	91	63	84	63	53	60	4
165	3	E	0	1	260	7	3	42	49	98	94	98	71	66	71	4
166	3	E	1	0	-	-	-	-	-	77	74	79	57	56	58	2
167	3	F	0	0	245	66	31	74	20	49	44	48	50	48	49	4
168	3	F	0	1	205	9	4	43	36	93	56	82	65	52	59	4
169	3	F	1	0	295	22	10	50	20	-	-	-	-	-	-	4
170	3	B	1	1	-	-	-	-	-	96	80	93	68	58	65	0
171	3	F	1	1	-	-	-	-	-	91	71	91	63	56	63	0

Table 32

Data for Students in Version 4 of the Material

						Error				SCAT (Perc.)			SCAT (T)			
Id.#	Version	Group	Sex	Grade	Time	R	%	T	Test	Verb.	Quant.	Tot.	Verb.	Quant.	Tot.	Status
172	4	A	0	0	200	8	7	46	46	98	92	98	71	64	71	4
173	4	A	1	0	290	23	21	64	22	-	01	01	-	27	27	4
174	4	A	0	1	-	-	-	-	-	39	60	48	47	53	49	2
175	4	A	1	1	235	4	4	41	16	58	13	39	52	39	47	4
176	4	B	1	0	-	-	-	-	-	27	04	14	44	32	39	3
177	4	B	0	1	195	5	14	54	20	93	95	97	65	66	69	4
178	4	B	1	1	375	3	12	52	33	75	56	70	57	52	55	4
179	4	B	0	0	145	0	0	36	33	93	88	94	65	62	66	4
180	4	C	1	0	290	1	0	49	38	89	98	97	62	71	69	4
181	4	C	1	1	135	6	6	44	53	98	96	99	71	68	73	4
182	4	C	0	0	190	9	7	71	26	95	98	98	66	71	71	4
183	4	D	0	1	-	-	-	-	-	49	48	49	50	49	50	2
184	4	D	1	1	315	4	4	41	52	80	94	91	58	66	63	4
185	4	D	0	0	235	7	6	45	34	98	97	99	71	69	73	4
186	4	D	1	0	265	2	2	39	20	61	52	59	53	51	52	4
187	4	E	1	1	-	-	-	-	-	36	04	19	46	32	41	2
188	4	E	0	0	225	5	14	54	31	96	86	95	68	61	66	4
189	4	E	1	0	-	-	-	-	-	-	-	-	-	-	-	2
190	4	E	1	1	-	-	-	-	-	98	99	99	71	73	73	2
191	4	F	1	1	-	-	-	-	-	49	03	25	50	31	43	3
192	4	F	0	0	270	24	22	65	39	58	56	59	52	52	52	4
193	4	F	1	0	-	-	-	-	-	42	22	64	48	42	54	2
194	4	C	0	1	-	-	-	-	-	72	83	81	56	60	59	1
195	4	F	0	1	-	-	-	-	-	49	71	61	50	56	53	0

Table 33

Data for Students in Version 5 of the Material

						Error				SCAT (Perc.)			SCAT (T)			
Id.#	Version	Group	Sex	Grade	Time	R	%	T	Test	Verb.	Quant.	Tot.	Verb.	Quant.	Tot.	Status
196	5	A	0	0	190	14	26	62	17	44	80	62	48	58	53	4
197	5	A	1	0	195	3	6	39	48	98	80	95	71	58	66	4
198	5	A	1	1	235	5	9	44	27	-	-	-	-	-	-	4
199	5	B	0	0	300	16	30	66	49	85	48	73	60	49	56	4
200	5	B	1	0	-	-	-	-	-	77	48	68	57	49	55	3
201	5	B	0	1	220	5	9	44	24	63	83	76	53	60	57	4
202	5	B	1	1	-	-	-	-	-	18	01	07	41	27	35	2
203	5	C	0	1	405	6	11	46	51	98	06	66	71	34	54	4
204	5	C	1	1	-	-	-	-	-	20	04	10	42	32	37	2
205	5	C	1	0	250	5	9	44	38	97	80	94	69	58	66	4
206	5	C	0	0	-	-	-	-	1	75	77	79	57	57	58	2
207	5	D	0	1	265	4	7	41	15	41	20	32	48	42	45	4
208	5	D	0	1	-	-	-	-	-	69	77	76	55	57	57	2
209	5	D	1	0	235	4	7	41	44	98	77	94	71	57	66	4
210	5	D	1	0	-	-	-	-	-	-	-	-	-	-	-	3
211	5	E	0	1	340	3	6	39	12	-	-	-	-	-	-	4
212	5	E	1	1	285	9	17	52	34	49	74	62	50	56	53	4
213	5	E	0	0	245	14	26	62	27	-	-	-	-	-	-	4
214	5	E	1	0	185	18	33	71	22	17	40	44	40	47	48	4
215	5	F	0	0	145	8	15	50	19	-	-	-	-	-	-	4
216	5	F	1	0	195	11	20	56	32	49	80	66	50	58	54	4
217	5	F	0	1	120	5	9	44	49	98	99	99	71	73	73	4
218	5	F	1	1	-	-	-	-	-	91	80	90	63	58	63	0

Table 34

Data for Students in Version 6 of the Material

						Error				SCAT (Perc.)			SCAT (T)			
Id.#	Version	Group	Sex	Grade	Time	R	%	T	Test	Verb.	Quant.	Tot.	Verb.	Quant.	Tot.	Status
219	6	A	1	0	-	-	-	-	-	41	20	32	48	42	45	3
220	6	A	0	0	-	-	-	-	-	87	60	79	61	53	58	2
221	6	A	1	1	250	3	11	47	42	-	-	-	-	-	-	4
222	6	A	0	1	-	-	-	-	-	72	96	90	56	68	63	2
223	6	B	0	0	-	-	-	-	-	-	01	01	-	27	27	3
224	6	B	1	1	300	6	22	54	42	87	52	76	61	51	57	4
225	6	B	0	1	235	18	67	81	9	-	-	-	-	-	-	4
226	6	B	1	0	360	5	19	52	21	24	52	33	43	51	46	4
227	6	C	0	0	225	3	11	47	48	99	94	98	73	66	71	4
228	6	C	1	0	325	2	7	45	46	66	74	73	54	56	56	4
229	6	D	1	1	290	2	7	45	46	96	77	92	68	57	64	4
230	6	D	0	1	-	-	-	-	-	-	-	-	-	-	-	2
231	6	D	1	0	250	1	4	43	44	97	80	94	69	58	66	4
232	6	D	0	0	-	-	-	-	-	39	40	39	47	47	47	2
233	6	E	1	1	230	1	4	43	50	89	86	91	62	61	63	4
234	6	E	1	0	245	2	7	45	37	93	36	75	65	46	57	4
235	6	F	0	1	-	-	-	-	-	93	48	79	65	49	58	2
236	6	F	1	0	235	6	22	54	34	63	52	61	53	51	53	4
237	6	F	0	0	175	2	7	45	20	39	44	40	47	48	47	4
238	6	F	1	1	-	-	-	-	-	27	28	25	44	44	43	2
239	6	C	1	1	-	-	-	-	-	55	13	37	51	39	47	1
240	6	C	0	1	-	-	-	-	-	39	94	70	47	66	55	1
241	6	E	0	1	-	-	-	-	-	77	99	95	57	73	66	1
242	6	E	1	0	-	-	-	-	-	-	-	-	-	-	-	0

Table 35

Data for Students in Version 7 of the Material[a]

							SCAT (Perc.)			SCAT (T)			
Id.#	Version	Group	Sex	Grade	Time	Test	Verb.	Quant.	Tot.	Verb.	Quant.	Tot.	Status
243	7	A	0	o	265	24	20	24	19	42	43	41	4
244	7	A	1	1	270	38	85	56	76	605	52	57	4
245	7	A	1	o	290	30	63	63	66	53	53	54	4
246	7	A	0	1	285	26	58	60	61	52	53	53	4
247	7	B	0	0	335	34	-	-	-	-	-	-	4
248	7	B	1	0	240	42	89	98	97	62	71	69	4
249	7	C	0	0	440	38	85	88	90	60	62	63	4
250	7	C	1	1	270	46	-	-	-	-	-	-	4
251	7	C	1	0	200	48	99	92	98	73	64	71	4
252	7	D	1	1	275	45	87	60	79	61	53	58	4
253	7	D	0	0	185	18	52	83	70	51	60	50	4
254	7	D	1	0	110	34	-	-	-	-	-	-	4
255	7	D	0	1	-	-	89	74	87	62	56	61	2
256	7	E	1	0	275	35	91	80	90	63	58	63	4
257	7	E	0	0	400	37	85	94	93	60	66	65	4
258	7	E	1	1	-	-	69	74	75	55	56	57	2
259	7	F	0	1	290	41	36	80	57	46	58	52	4
260	7	F	0	0	305	45	-	-	-	-	-	-	4
261	7	F	1	1	260	37	58	88	76	52	62	57	4
262	7	C	0	1	275	22	34	40	35	46	47	46	4
263	7	B	1	1	-	-	-	-	-	-	-	-	1
264	7	B	0	1	-	-	96	99	99	68	73	73	0
265	7	E	0	1	-	-	-	-	-	-	-	-	0
266	7	F	1	0	-	-	93	32	73	65	45	56	1

---

[a] Since Version 7 did not require responding there are no error data for this version.

REFERENCES

Briggs, L. J., Learner variables and educational media. Rev. educ. Res., 1968, 38, 160-176.

Briggs, L. J., Goldbeck, R. A., Campbell, V., & Nichols, D. C., Experimental results regarding form of response, size of step,and individual differences in automated programs. In J. E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: Wiley, 1962. Pp. 86-98.

Cochran, W. G. & Cox, Gertrude M., Experimental designs, (2nd Ed.) New York: Wiley, 1957.

Coulson, J. E. & Silberman, H. F., Effects of three variables in a teaching machine. J. educ. Psychol., 1960, 51, 135-143.

Coulson, J. E. & Silberman, H. F., Results of initial experiments in automated teaching. Cited in L. M. Stolurow, Teaching by machine, Cooperative Res. Monogr. No. 6, Washington, D.C.: U. S. Government Printing Office, 1961.

Crowder, N. A., On the differences between linear and intrinsic programming. Phi Delta Kappan, 1963, 44, 250-254.

Deterline, W. A., Practical problems in program production. In P.C. Lange (Ed.), Programmed Instruction. Yearb. nat. Soc. Stud. Educ., 1967, 66, Part II, Pp. 178-216.

Ebel, R. L., Measuring educational achievement. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Edwards, A. L., Experimental design in psychological research. (2nd Ed.) New York: Holt, Rinehart & Winston, 1960.

Engelhart, M. D., A comparison of several item discrimination indices. J. educ. Measmt., 1965, 2, 69-76.

Evans, J. L., Programming in mathematics and logic. In R. Glaser (Ed.), Teaching machines and programmed learning II: Data and directions. Washington, D.C. : NEA, 1965, Pp. 371-440.

Flynn, J. M., Research on computers and instruction: A critique of Schurdak's article. Amer. educ. res. J. , 1967, 4, 329-336.

Fry, E. B., Teaching machine programming. New York: McGraw-Hill, 1963.

Green, R. F., Cooperative school and college ability tests. In O.K. Buros (Ed.) The sixth mental Measurement Yearbook, Highland Park, N.J.: The Gryphon Press, 1965

Hays, W. L. Statistics for psychologists, New York: Holt, Rinehart & Winston, 1963.

Holland, J. G. and Skinner, B. F. The analysis of behavior. New York: McGraw-Hill, 1961.

Kapel, D. E. An analysis of the effectiveness and efficiency of two modes of presentation and two step sizes, and subsequent effects of determinant variables in programmed instruction. J. prog. Instruc., 1965, 3 (2), 21-34.

Krumboltz, J. D. & Kiesler, C. A. The partial reinforcement paradigm and programmed instruction. J. prog. Instruc., 1965, 3 (2), 9-14.

Licklider, J. C. R. Preliminary experiments in computer-aided teaching. In J. E. Coulson (Ed.), Programmed learning and computer-based instruction. New York: Wiley, 1962. Pp. 217-239.

Lumsdaine, A. A. Some theoretical and practical problems in programmed instruction. In J. E. Coulson (Ed.), Programmed Learning and computer-based instruction. New York: Wiley, 1962. Pp. 134-151.

Pressey, S. L. Some perspectives and major problems regarding teaching machines. In A. A. Lumsdaine and R. Glaser (Eds.) Teaching machines and programmed learning: a source book. Washington, D. C.: NEA, 1960. Pp. 497-505.

Scandura, J. M. Teaching - technology or theory. Amer. educ. Res. J., 1966, 3, 139-146.

Schramm, W. The research on programmed instruction: an annotated bibliography. U. S. O. E. Bulletin No. 35, 1964.

Shay, C. B. Relationship of intelligence to step size on a teaching machine program. J. educ. Psychol., 1961, 52, 98-103.

Silberman, H. F. In J. E. Coulson (Ed.), Programmed learning and computer-based instruction. New York: Wiley, 1962, Pp. 13-24.

Skinner, B. F. Teaching machines. Sci., 1958, 128, 969-977.

Smith, W. & Moore, J. W. Size-of-step and achievement in programmed spelling. Psychol. Rep., 1962, 10, 287-294.

Snedecor, G. W. Statistical methods, Ames. Iowa: Iowa State Univ. Press, 1956.

Stolurow, L. M. Programmed instruction for the mentally retarded. Rev. educ. Res., 1963, 33, 126-136.

Tuel, J. K. The relationship of intelligence and achievement variables in programmed instruction. Calif. J. Educ. Res., 1966, 17, 68-72.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

BIOGRAPHICAL SKETCH

John Michael Flynn was born on January 19, 1940 in New Castle, Pennsylvania and he attended public schools in East Palestine, Unity, and Columbiana, Ohio. He graduated from South Dade High School in Homestead, Florida in 1958. He attended the University of Florida where he was active in student government and other extra curricular activities. After being out of school for several years, he returned and received his Bachelor of Science in Education degree with a major in mathematics education and a minor in psychology in 1965. He received his Master of Education degree at the University of Florida the following year in educational psychology.

While at the University of Florida, he served as a research assistant and interim instructor, and was a recipient of a National Defense Education Act fellowship. During his last year in graduate school, he held a position as a full time instructor at Nova University in Fort Lauderdale, Florida. He is a member of the American Educational Research Association, the Association for Computing Machinery, the Association of Educational Data Systems, Phi Delta Kappa (honorary), and Kappa Delta Phi (honorary). He is married to the former Bonnita Maureen Smalt of Lakeland, Florida and has two children.

This dissertation was prepared under the direction of the chairman of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Education and to the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Education.

August, 1968

B Sharp by M C Baker
Dean, College of Education

Dean, Graduate School

Supervisory Committee:

Wilson H. Guertin
Chairman

D. E. Small